

RO/KR 29. 04. 2004

대한민국 특허청  
KOREAN INTELLECTUAL  
PROPERTY OFFICE

RECD 18 MAY 2004

WIPO

PCT

별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto  
is a true copy from the records of the Korean Intellectual  
Property Office.

출원번호 : 10-2003-0020023  
Application Number

**PRIORITY DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)

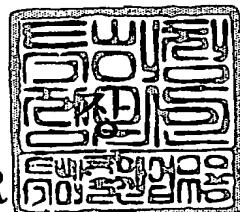
출원년월일 : 2003년 03월 31일  
Date of Application MAR 31, 2003

출원인 : 주식회사 알진텍 외 2명  
Applicant(s) ALGENTECH, et al.

2004년 04월 29일

특허청

COMMISSIONER



## 【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【제출일자】	2003.03.31
【발명의 명칭】	카로티노이드계 색소 생합성에 관여하는 유전자
【발명의 영문명칭】	Genes involved in the biosynthesis of carotenoids
【출원인】	
【성명】	김영태
【출원인코드】	4-1998-039808-4
【출원인】	
【성명】	이재형
【출원인코드】	4-2003-011731-1
【출원인】	
【명칭】	주식회사 알진텍
【출원인코드】	1-2001-008962-2
【대리인】	
【성명】	이원희
【대리인코드】	9-1998-000385-9
【포괄위임등록번호】	2003-019988-1
【포괄위임등록번호】	2003-019853-9
【포괄위임등록번호】	2003-004687-1
【발명자】	
【성명】	김영태
【출원인코드】	4-1998-039808-4
【발명자】	
【성명】	이재형
【출원인코드】	4-2003-011731-1
【심사청구】	청구
【미생물기탁】	
【기탁기관명】	한국미생물보존센터 (KCCM)
【수탁번호】	KCCM-10460
【수탁일자】	2003.01.24

## 【핵산영기 및 아미노산 서열목록】

【서열개수】	15	
【서열목록의 전자파일】	첨부	
【취지】	특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사 를 청구합니다. 대리인 이원희 (인)	
【수수료】		
【기본출원료】	20 면	29,000 원
【가산출원료】	33 면	33,000 원
【우선권주장료】	0 건	0 원
【심사청구료】	16 항	621,000 원
【합계】	683,000 원	
【감면사유】	소기업 (70%감면)	
【감면후 수수료】	204,900 원	
【첨부서류】	1. 요약서·명세서(도면)_1통 2. 미생물기탁증명서_1통 3. 소기업임 을 증명하는 서류_1통	

**【요약서】****【요약】**

본 발명은 카로티노이드 생합성에 관여하는 유전자에 관한 것으로서, 보다 상세하게는 카로티노이드 생합성에 필요한 단백질을 코딩하는 서열번호 2, 서열번호 4, 서열번호 6, 서열번호 8, 서열번호 10 및 서열번호 12로 기재되는 유전자를 포함하는 카로티노이드 생합성 유전자에 관한 것이다. 본 발명에 따른 카로티노이드 생합성 유전자는 카로티노이드의 대량 생산에 매우 유용하게 사용될 수 있다.

**【대표도】**

도 8

**【색인어】**

카로티노이드, 아스타잔틴

## 【명세서】

## 【발명의 명칭】

카로티노이드계 색소 생합성에 관여하는 유전자{Genes involved in the biosynthesis of carotenoids}

## 【도면의 간단한 설명】

도 1은 카로티노이드의 생합성 경로를 도시한 것이다.

도 2는 파라코커스 해운대시스로부터 분리된 DNA 단편에 존재하는 첫 번째 ORF(open reading frame)의 아미노산 서열과 알칼리제네스 속 미생물과 브라디리조비움 속 미생물로부터 분리된  $\beta$ -카로텐 케토레이즈( $\beta$ -carotene ketolase; crtW)의 아미노산 서열을 비교한 결과이다.

P. haeundaesis: 파라코커스 해운대시스

Alcaligenes\_sp: 알칼리제네스 속

Bradyrhizobium\_sp: 브라디리조비움 속

Consensus: 일치하는 아미노산 서열

도 3은 파라코커스 해운대시스로부터 분리된 DNA 단편에 존재하는 두 번째 ORF의 아미노산 서열과 알칼리제네스 속 미생물로부터  $\beta$ -카로텐 하이드록실레이즈( $\beta$ -carotene hydroxylase; crtZ)의 아미노산 서열을 비교한 결과이다.

P. haeundaesis: 파라코커스 해운대시스

Alcaligenes\_sp: 알칼리제네스 속

Consensus: 일치하는 아미노산 서열

도 4는 파라코커스 해운대시스로부터 분리된 DNA 단편에 존재하는 세 번째 ORF의 아미노산 서열과 플라보박테리움 속 미생물로부터 분리된 리코펜 싸이클레이즈(lycopene cyclase; crtY)의 아미노산 서열을 비교한 결과이다.

*P. haeundaesis*: 파라코커스 해운대시스

*Flavobacterium\_sp*: 플라보박테리움 속

Consensus: 일치하는 아미노산 서열

도 5는 파라코커스 해운대시스로부터 분리된 DNA 단편에 존재하는 네 번째 ORF의 아미노산 서열과 플라보박테리움 속 미생물로부터 분리된 피토엔 디세튜레이즈(phytoene desaturase; crtI)의 아미노산 서열을 비교한 결과이다.

*P. haeundaesis*: 파라코커스 해운대시스

*Flavobacterium\_sp*: 플라보박테리움 속

Consensus: 일치하는 아미노산 서열

도 6은 파라코커스 해운대시스로부터 분리된 DNA 단편에 존재하는 다섯 번째 ORF의 아미노산 서열과 플라보박테리움 속 미생물로부터 분리된 피토엔 신테이즈(phytoene synthase; crtB)의 아미노산 서열을 비교한 결과이다.

*P. haeundaesis*: 파라코커스 해운대시스

*Flavobacterium\_sp*: 플라보박테리움 속

Consensus: 일치하는 아미노산 서열

도 7은 파라코커스 해운대시스로부터 분리된 DNA 단편에 존재하는 여섯 번째 ORF의 아미노산 서열과 플라보박테리움 속 미생물로부터 분리된 제라닐제라닐 피로포스페이트 신테이즈(geranylgeranyl pyrophosphate synthase; crtE)의 아미노산 서열을 비교한 결과이다.

*P. haeundaesis*: 파라코커스 해운대시스

*Flavobacterium\_sp*: 플라보박테리움 속

Consensus: 일치하는 아미노산 서열

도 8은 파라코커스 해운대시스로부터 분리된 *crt* 유전자의 구성(organization)을 도시화한 것이다.

도 9는 본 발명의 pCR-XL-TOPO-crtfull 벡터의 개열지도를 나타낸 것이다.

도 10은 본 발명의 *crt* 유전자가 도입된 형질전환체의 배양액으로부터 추출된 메탄을 추출액의 흡광도 변화(190-890 nm:A 및 350-550 nm :B)를 스캐닝한 결과를 도시한 것으로서, 450 nm의 피크는  $\beta$ -카로텐의 고유 피크이고, 470 nm의 피크는 아스타잔틴의 고유 피크이다.

### 【발명의 상세한 설명】

### 【발명의 목적】

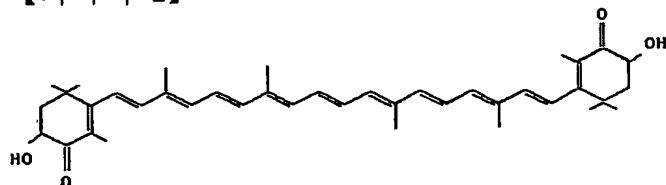
### 【발명이 속하는 기술분야 및 그 분야의 종래기술】

<30> 본 발명은 카로티노이드 생합성에 관여하는 유전자에 관한 것으로서, 보다 상세하게는 카로티노이드 생합성에 필요한 서열번호 2, 서열번호 4, 서열번호 6, 서열번호 8, 서열번호 10 및 서열번호 12의 유전자를 포함하는 카로티노이드 생합성 유전자에 관한 것이다.

<31> 카로티노이드(carotenoids)는 항산화 활성을 가지고 있는  $C_{40}$  이소프레노이드 화합물(isoprenoid compounds)로서, 자연계에 널리 분포되어 있는 한 군의 색소의 총칭을 말한다. 현재까지 알려진 카로티노이드는 6백여 종에 이르며, 이들은 각각 다른 형태로 존재한다. 카로티노이드는 분자 구조에 따라 노란색, 적색, 주홍색, 주황색 등 여러 색으로 존재한다. 그 예로는  $\beta$ -카로텐( $\beta$ -carotene; 당근의 주황색 색소), 리코펜(lycopene; 토마토의 적색 색소), 푸코잔틴(fucoxanthin; 해조류의 황갈색 또는 갈색 색소) 등이 있다. 카로티노이드는 인체 내에서 비타민 A의 전구체로서의 역할을 하며, 산화 방지효과와 유해산소 소거작용, 암세포의 종식 억제작용 및 발암 억제작용이 강하여 순환기 질환, 암 및 성인병 등을 예방하는 효능을 가진 것으로 알려져 있다. 최근에는 카로티노이드가 직접적인 자외선에 의한 신체의 면역기능을 향상시켜 자외선 노출로부터 피부의 손상을 줄여주거나 멜라닌 생성을 억제함이 밝혀지면서 유럽이나 미국에서 미용 소재로 각광을 받기 시작했다. 현재 카로티노이드는 건강식품 소재(영양 보충제), 인간을 대상으로 한 약학적 제제와 식품 착색제 또는 동물용 사료의 색소 등으로 사용되고 있다.

<32> 카로티노이드 중에서도 하기 화학식 1로 표시되는 구조를 가지는 아스타잔틴(3, 3'-dihydroxy- $\beta$ , $\beta$ -carotene-4, 4'-dione)은 천연적으로 생산되는 주홍색 또는 밝은 오렌지색의 색소 물질이다.

<33> 【화학식 1】



<34> 아스타잔틴은 주로 새우, 붉은 도미류(red seabream), 연어(salmon) 및 바다가재(lobsters) 등과 같은 해양 동물 조직에 존재하는 것으로 알려져 있다(Fujita *et al.*, *Nippon Suisan Gakkaishi*, 49: 1855-1869, 1983; Johnson, E. A., *Crit. Rev. Biotechnol.*, 11: 297-326, 1991; Nelis *et al.*, *J. Appl. Bacteriol.*, 70: 181-191, 1991). 아스타잔틴은 정상적인 호기적 대사과정 중 활성산소가 세포 내 DNA, 단백질, 지질 등을 손상시키는 것과 더불어 세포와 조직의 노화 및 발암을 유발하는 반응을 억제할 뿐만 아니라, 유리 라디칼(hydroxy or peroxy radicals)의 생성을 억제하는 작용을 한다(Palozza *et al.*, *Arch. Biochem. Biophys.*, 297: 291-295, 1992; Shimidzu *et al.*, *Fish Sci.*, 62: 134-137, 1996). 또한, 아스타잔틴은 면역 조절 활성(immune modulatory activity) 및 심장 조절 효과(cardioprotective effect)를 갖는 것으로 알려졌다(Jyonuchi *et al.*, *Nutr. Cancer.*, 19: 269-280, 1993). 특히, 아스타잔틴의 항산화력은 다른 카로티노이드의 10배 이상이고,  $\alpha$ -토코페롤(tocopherol)의 100배 이상인 것으로 알려져 있다. 또한, 현재까지 아스타잔틴의 독성(toxicity)에 대해서는 보고된 바 없다. 따라서, 아스타잔틴은 신경질환(neurodegenerative diseases), 암(cancer), 면역질환(immune disorders), 심장혈관질환(cardiovascular diseases) 등을 포함하는 다양한 질환의 치료 및 예방에 이용되고 있고, 또한 이에 대한 연구가 계속 되고 있다(Beal, H. F., *The Neuroscientist*, 3: 21-27, 1991; Chew *et al.*, *Anticancer Res.*, 19: 1849-1853, 1999; Murillo E., *Arch. Latinoam. Nutr.*, 42: 409-413, 1992). 또한, 아스타잔틴은 산업적으로 색소증진 물질로 이용되고 있으며, 우리나라에서는 '파파야 색소'라는 명칭으로 식품첨가물로 등록되어 있다. 이러한 이유로 아스타잔틴의 수요는 매년 15% 이상의 소비성장을 보이고 있으며, 이에 따라 아스타잔틴의 중요성이 대두되고 있다.

<35> 최근에 스위스의 호프만-라로체(F. Hoffman-LaRoche) 사에서 아스타잔틴의 화학 합성법을 개발하였으나, 화학합성으로 제조된 아스타잔틴은 천연 아스타잔틴에 비해 낮은 생체 흡수율을 보이고, 식품 첨가제로서의 안전성에 문제가 되고 있어 유럽의 일부 국가에서만 사용이 허가된 상태이다. 따라서, 천연 아스타잔틴의 합성법에 관심이 집중되고 있으며, 아스타잔틴을 생산하는 미생물을 이용한 아스타잔틴의 제조에 상업적 관심이 일고 있다. 현재 아스타잔틴을 생산하는 것으로 알려진 미생물로는 효모인 파피아 로도지마(*Phaffia rhodozyma*; Miller et al., *Int. J. Syst. Bacteriol.*, 48: 529-536, 1976), 녹색조류인 해마토코커스 플루비알리스(*Haematococcus pluvialis*; Bubrick, *Bioresour Technol.*, 38: 237-239, 1991), 그람-양성균 브레비박테리움 103(*Brevibacterium 103*; Lizuka & Nishimura, *J. Gen. Appl. Microbiol.*, 15: 127-134, 1969), 그람-음성균 아그로박테리움 아우란티아쿰(*Agrobacterium aurantiacum*; Yokoyama et al., *Biosci. Biotechnol. Biochem.*, 58: 1842-1844, 1994), 파라코커스 마르쿠시아이(*Paracoccus marcusii*; Harker et al., *Int. J. Syst. Bacteriol.*, 48: 543-548, 1998) 및 파라코커스 카로티니파시엔스(*Paracoccus carotinifaciens*; Tsubokura et al., *Int. J. Syst. Bacteriol.*, 49: 277-282, 1999) 등이 알려져 있다.

<36> 한편, 최근 6년 동안 카로티노이드 생합성에 관여하는 효소들을 코딩하는 유전자들에 대해 연구되어 왔다. 그 결과, 많은 카로티노이드 생합성 유전자들이 다양한 미생물로부터 클로닝되었으며, 이들의 기능이 규명되었다(Armstrong, G. A., *J. Bacteriol.*, 176: 4795-4802, 1994; Sandmann, G., *Eur. J. Biochem.*, 223: 7-24, 1994; Wieland, B., *J. Bacteriol.*, 176: 7719-7726, 1994). 카로티노이드 생합성 경로는 일반적인 이소프레노이드(isoprenoid) 경로의 중요한 중간 생성물인 FPP(farnesyl pyrophosphate)로부터 파생한다. 도 1을 참조하면, FPP

와 IPP(isopentenyl pyrophosphate)는 *crtE*에 의해 암호화되는 제라닐제라닐 피로포스페이트 신테이즈(geranylgeranyl pyrophosphate synthase)에 의해 GGPP(geranylgeranyl pyrophosphate)로 된다. 이후, GGPP는 *crtB*에 의해 암호화되는 피토엔 신테이즈(phytoene synthase), *crtI*에 의해 암호화되는 피토엔 디세튜레이즈(phytoene desaturase), 그리고 *crtY*에 의해 암호화되는 리코펜 싸이클레이즈(lycopene cyclase)에 의해 일련의 반응을 거쳐  $\beta$ -카로텐로 전환된다.  $\beta$ -카로텐은 *crtW*에 의해 암호화되는  $\beta$ -카로텐 케토레이즈( $\beta$ -carotene ketolase)와 *crtZ*에 의해 암호화되는  $\beta$ -카로텐 하이드록실레이즈( $\beta$ -carotene hydroxylase)에 의해 일련의 반응을 거쳐 아스타잔틴으로 전환된다.

<37> 카로티노이드 생합성에 관여하는 *crt* 유전자(carotenogenic gene)의 염기서열, 이의 구성(organization) 및 단백질의 특성이 로도박터 캡슐라터스(*Rhodobacter capsulatus*)(Armstrong et al., *Mol. Gen. Genet.*, 216: 254-268, 1989)와 에르위니아 헤르비콜라(*Erwinia herbicola*)(Sandmann et al., *FEMS Microbiol. Lett.*, 71: 77-82, 1990; Hundle et al., *Photochem. Photobiol.*, 54: 89-93, 1991) 및 에르위니아 우레도보라(*Erwinia uredovora*)(Misawa et al., *J. Bacteriol.*, 172: 6704-6712, 1990)에서 규명된 바 있다. 또한, *crtB*, *crtI*, *crtY*, *crtW* 및 *crtZ*로 이루어진 카로티노이드 생합성 *crt* 유전자가 해양미생물인 아그로박테리움 아우란티아컴으로부터 분리된 바 있으며(Norihiko et al., *J. Bacteriol.*, 177(22): 6575-6584, 1995), GGPP로부터  $\beta$ -카로텐까지의 반응을 촉매하는 효소들을 코딩하는 3개의 유전자들(*crtB*, *crtI* 및 *crtY*)과 이들이 도입된 파파아 로도지마에 대해 보고된 바 있다(WO 97/23633).

<38> 이에, 본 발명자들은 아스타잔틴을 생산하는 파라코커스 해운대시스로부터 카로티노이드 생합성에 관여하는 유전자들을 분리하기 위해 연구를 거듭하던 중, *crtE*, *crtB*, *crtI*, *crtY*, *crtW* 및 *crtZ* 유전자, 그리고 이들을 포함하는 *crt* 유전자를 클로닝하여 그의 염기서열을 규명하고, 상기 *crt* 유전자를 이용하여 카로티노이드를 생산하지 않는 미생물에서 카로티노이드를 생산할 수 있음을 규명함으로써 본 발명을 완성하였다.

#### 【발명이 이루고자 하는 기술적 과제】

<39> 본 발명의 목적은 카로티노이드 생합성에 필요한 단백질을 코딩하는 유전자를 제공하는 것이다.

<40> 본 발명의 또 다른 목적은 상기 유전자를 포함하는 재조합 벡터를 제공하는 것이다.

<41> 나아가, 본 발명의 또 다른 목적은 상기 유전자를 이용하여 카로티노이드를 생산하는 방법을 제공하는 것이다.

#### 【발명의 구성】

<42> 상기와 같은 목적을 달성하기 위하여, 본 발명은 카로티노이드 생합성에 필요한 단백질을 코딩하는 서열번호 2, 서열번호 4, 서열번호 6, 서열번호 8, 서열번호 10 및 서열번호 12로 구성된 군으로부터 선택되는 염기서열을 갖는 유전자를 제공한다.

<43> 또한, 본 발명은 상기 유전자를 포함하는 카로티노이드 생합성 유전자를 제공한다. 구체적으로 본 발명은 서열번호 1로 기재되는 염기서열을 갖는 *crt* 유전자를 제공한다.

<44> 또한, 본 발명은 상기 카로티노이드 생합성 유전자를 포함하는 재조합 벡터를 제공한다.

<45> 또한, 본 발명은 상기 카로티노이드 생합성 유전자를 이용하여 카로티노이드를 생산하는 방법을 제공한다.

<46> 이하, 본 발명을 상세히 설명한다.

<47> 본 발명은 카로티노이드 생합성에 필요한 단백질을 코딩하는 서열번호 2, 서열번호 4, 서열번호 6, 서열번호 8, 서열번호 10 및 서열번호 12로 이루어진 군에서 선택되는 염기서열을 갖는 유전자를 제공한다.

<48> 상기 유전자는 아스타잔틴을 생산하는 파라코커스 해운대시스(기탁번호:KCCM-10460)로부터 분리된 유전자이다.

<49> 본 발명의 카로티노이드 생합성에 필요한 단백질을 코딩하는 유전자는 서열번호 2, 서열번호 4, 서열번호 6, 서열번호 8, 서열번호 10 및 서열번호 12로 구성된 군으로부터 선택되는 유전자인 것이 바람직하다. 상기 서열번호 2, 서열번호 4, 서열번호 6, 서열번호 8, 서열번호 10 및 서열번호 12로 기재되는 염기서열은 각각 서열번호 3, 서열번호 5, 서열번호 7, 서열번호 9, 서열번호 11 및 서열번호 13으로 기재되는 아미노산을 암호화하는 서열이다. 본 발명의 6개의 유전자 및 이로부터 코딩되는 단백질에 대하여 하기 표 1에 기재하였다.

<50>

【표 1】

유전자	유전자명	단백질	단백질의 아미노산 서열
서열번호 2	<i>crtW</i>	β-카로텐 케토레이즈 (β-carotene ketolase)	서열번호 3
서열번호 4	<i>crtZ</i>	β-카로텐 하이드록실레이즈 (β-carotene hydroxylase)	서열번호 5
서열번호 6	<i>crtY</i>	리코펜 싸이클레이즈(lycopene cyclase)	서열번호 7
서열번호 8	<i>crtI</i>	피토엔 디세튜레이즈 (phytoene desaturase)	서열번호 9
서열번호 10	<i>crtB</i>	피토엔 신테이즈(phytoene synthase)	서열번호 11
서열번호 12	<i>crtE</i>	제라닐제라닐 피로포스페이트 신테이즈 (geranylgeranyl pyrophosphate synthase)	서열번호 13

<51> 본 발명에 따라 제공되는 유전자들은 다양한 숙주세포에 도입되어 카로티노이드를 생산하는데 유용하게 사용될 수 있다. 상기 유전자들은 단독으로 사용될 수도 있으며, 2개 이상이 함께 사용될 수도 있다. 예컨대, 리코펜 싸이클레이즈를 코딩하는 서열번호 6으로 기재되는 유전자는 *crtE*, *crtB* 및 *crtI*만을 가지고 있는 미생물에 도입되어 β-카로텐을 생산하는데 사용될 수 있으며, β-카로텐 케토레이즈 및 β-카로텐 하이드록실레이즈를 각각 코딩하는 서열 번호 2 및 서열번호 4로 기재되는 유전자는 β-카로텐을 생산하는 미생물(예: 파피라 토도지마 ATCC96815)에 도입되어 아스타잔틴을 생산하는데 사용될 수 있다.

<52> 또한, 본 발명은 상기 유전자들을 모두 포함하는 카로티노이드 생합성 유전자를 제공한다.

<53> 상기 카로티노이드 생합성 유전자는 서열번호 1로 기재되는 염기서열을 갖는 것이 바람직하다. 본 발명의 카로티노이드 생합성 유전자(carotenoid synthesis gene, 이하 'crt 유전자'라 약칭함)는 FPP(farnensyl pyrophosphate)로부터 아스타잔틴을 생산하는 과정에 관여하는

모든 카로티노이드 생합성 유전자들을 포함한다. 본 발명의 *crt* 유전자의 구성을 도 8에 도시하였다. 도 8에서 보는 바와 같이, 본 발명의 *crt* 유전자는 6,223 bp의 크기이며, 그 내부에는 5'→3' 방향으로 *crtW*, *crtZ*, *crtY*, *crtI*, *crtB* 및 *crtE* 유전자가 순서대로 위치하고 있다. 또한, 그 내부에는 *Kpn* I, *Sma* I, *Xma* I, *Cla* I, *Hind* III 및 *Bam* H I 인식서열이 하나씩 존재하고 있다. *crtW*, *crtZ*, *crtY*, *crtI* 및 *crtB* 유전자들은 종결코돈(stop codon)과 그 다음 유전자의 시작코돈(start codon)이 중첩(overlap)되어 존재한다. 특히, *crtE* 유전자는 상보적 가닥(complementary strand)의 방향으로 존재한다.

<54> 또한, 본 발명은 상기 카로티노이드 생합성 유전자를 포함하는 재조합 벡터를 제공한다.

<55> 본 발명의 재조합 벡터는 *crt* 유전자를 기본 벡터에 삽입하여 제조한 것이다. 본 발명에서 사용될 수 있는 기본 벡터는 유전자의 클로닝 또는 발현에 일반적으로 사용되는 벡터라면 제한없이 사용될 수 있다. 또한, 벡터는 숙주세포에 따라 달라질 수 있다. 즉, 숙주세포로 대장균을 사용하는 경우에는 대장균의 복제기원을 가지고 있는 대장균용 벡터를 사용하는 것이 바람직하고, 효모를 사용하는 경우에는 효모의 복제기원을 가지고 있는 효모용 벡터를 사용하는 것이 바람직하다. 또한, 대장균과 효모의 복제기원을 둘 다 가지고 있는 셔틀 벡터(shuttle vector)를 사용할 수도 있다. 본 발명의 바람직한 실시예에서는 pCR-XL-TOPO 벡터를 사용하여 *crt* 유전자를 포함하는 재조합 벡터를 제조하였으며, 이를 'pCR-XL-TOPO crtfull'이라 명명하였다.

<56> 또한, 본 발명은 상기 카로티노이드 생합성 유전자를 포함하는 재조합 벡터를 숙주세포에 형질도입한 균주를 제공한다.

<57> 상기에서 숙주세포는 대장균 또는 효모를 사용할 수 있으며, 대장균은 XL1-Blue, TOPO, BL21(DE3) 코돈 플러스(codon plus), DH1 및 DH5  $\alpha$  균주로 구성된 균으로부터 선택되는 것이 바람직하나 반드시 이에 한정되는 것은 아니다. 본 발명의 바람직한 실시예에서는 서열번호 1로 기재되는 *crt* 유전자를 포함하는 재조합 벡터 pCR-XL-TOP0 crtfull을 대장균인 BL21(DE3) 코돈 플러스에 형질도입한 균주를 제조하였다.

<58> 또한, 본 발명은 카로티노이드 생합성 유전자를 이용하여 카로티노이드를 생산하는 방법을 제공한다.

<59> 본 발명의 카로티노이드 생산 방법은

<60> 1) 서열번호 1의 *crt* 유전자를 클로닝하는 단계;

<61> 2) 단계 1의 유전자를 삽입한 재조합 벡터를 제조하는 단계;

<62> 3) 단계 2의 재조합 벡터를 숙주세포에 형질도입하는 단계;

<63> 4) 단계 3의 형질도입된 균주를 배양하여 그 배양액으로부터 카로티노이드를 회수하는 단계를 포함한다.

<64> 상기에서 숙주세포로는 대장균을 사용할 수 있다. 이 때 대장균으로는 일반적인 형질전환에 사용되는 것이라면 제한없이 사용될 수 있으나, XL1-Blue, TOPO, BL21(DE3) 코돈 플러스(codon plus), DH1 및 DH5  $\alpha$ 로 구성된 균으로부터 선택되는 것을 사용하는 것이 바람직하다.

본 발명의 바람직한 실시예에서는 BL21(DE3) 코돈 플러스를 사용하였다. 또한, 본 발명의 방법에 사용될 수 있는 숙주세포는 대장균에만 제한되는 것은 아니며, 효모를 사용할 수 있다.

<65> 상기에서 배양은 단계 1 내지 단계 3을 통해 제조된 균주를 생육가능한 배지에서 1차 배양하고 배양액으로부터 균체를 회수한 후, 다시 상기 균체에 유기용매를 첨가하여 4°C에서 하룻밤 동안 2차 배양하는 것이 바람직하다. 이 때 배양액에 카로티노이드의 생성을 유도하는 유도제(inducer), 예컨대 IPTG(isopropyl-beta-D-thiogalactopyranoside)를 추가로 첨가할 수 있으며, 또한 카로티노이드 기질, 예컨대 FPP(farnesyl pyrophosphate), GGPP(geranylgeranyl diphosphate) 또는 GPP(geranyl pyrophosphate)를 추가로 첨가할 수 있다. 상기 배양액으로부터 카로티노이드를 용출하기 위한 유기용매로는 메탄올, 아세톤 또는 에틸에테르를 사용할 수 있으며, 보다 바람직하게는 메탄올을 사용하는 것이 바람직하다. 또한, 배양액으로부터 카로티노이드의 회수는 HPLC(high performance liquid chromatography) 또는 TLC(thin-layer chromatography)를 이용하여 당업계에 공지된 방법에 따라 수행할 수 있다.

<66> 본 발명의 바람직한 실시예에서는 상기 단계를 통해 *crt* 유전자를 포함하는 형질전환 균주로부터 생산된 아스타잔틴의 양을 측정한 결과, 110  $\mu\text{g/g}$ (dry weight)를 생산하여 아스타잔틴 생산균주인 파라코커스 해운대시스가 생산하는 아스타잔틴의 양(25  $\mu\text{g/g}$ (dry weight)) 보다 훨씬 뛰어남을 확인하였다. 따라서, 본 발명의 아스타잔틴을 생산하는 방법은 아스타잔틴을 생산하지 않는 균주에서도 카로티노이드 생합성 유전자를 이용하여 대량으로 아스타잔틴을 생산함을 알 수 있어, 아스타잔틴을 이용한 식품첨가물로써의 색소 생산, 의약품의 제조 등에 유용하게 사용될 수 있다.

<67> 본 발명의 일 실시예에서는 카로티노이드 생합성에 필요한 단백질을 코딩하는 유전자들을 클로닝하기 위하여 파라코커스 해운대시스로부터 게노믹 DNA 라이브러리(genomic DNA library)를 구축하였다. 게노믹 DNA 라이브러리는 당업계에게 공지된 통상의 방법에 따라 구축할 수 있으며, 구체적으로 본 발명에서는 코스미드 벡터(cosmid vector)를 이용하여 게노믹 DNA 라이브러리를 구축하였다.

<68> 본 발명의 다른 실시예에서는 '색 상보화(color complementation)' 방법을 이용하여 게노믹 DNA 라이브러리로부터 카로티노이드 생합성에 관여하는 유전자들을 클로닝하였다. 카로티노이드를 생산하지 않는 미생물(예: 대장균)은 카로티노이드를 생산하는 미생물(예: 본 발명의 파라코커스 해운대시스)로부터 클로닝된 카로티노이드 생합성 관련 유전자들에 의해 형질전환됨으로써 카로티노이드를 생산할 수 있는 능력을 갖게 된다. 예컨대, *crtE*, *crtB*, *crtI* 및 *crtY* 유전자가 도입된 대장균은  $\beta$ -카로텐을 생산하고, 이에 따라 세포는 노란색을 띤다. 또한, *crtE*, *crtB*, *crtI*, *crtY*, *crtW* 및 *crtZ* 유전자가 모두 도입된 대장균은 아스타잔틴을 생산하고, 이에 따라 세포는 오렌지색을 띤다. 따라서, 본 발명자들은 파라코커스 해운대시스의 게노믹 DNA 라이브러리를 카로티노이드의 공통된 기질인 FPP(farnesyl pyrophosphate)를 첨가한 배지에서 배양하였다. 그 결과, 오렌지색을 띠는 13개의 콜로니를 선발할 수 있었다. 이후, 각 콜로니로부터 코스미드 벡터를 분리하여 벡터 내부에 삽입된 DNA 단편(insert)의 염기서열을 결정하였다. 그 결과, 가장 작은 크기의 DNA 단편은 6,223 bp의 크기임을 확인하였다. 상기 DNA 단편의 염기서열을 서열번호 1로 기재하였다.

<69> 본 발명의 또 다른 실시예에서는 카로티노이드를 생산하는 유전자가 존재하는 것으로 추정되는 6,223bp의 DNA 단편의 염기서열을 분석하여 6개의 ORF를 찾아내었다. 이들을 NCBI GenBank 상에서 분석한 결과, 각 ORF로부터 유추되는 아미노산 서열은 FPP로부터 아스타잔틴을

생성하는 반응에 관여하는 6개의 효소들의 아미노산 서열과 매우 높은 상동성을 나타내었다(도 2 내지 도 7 참조). 이 결과로부터 본 발명자들은 본 발명에서 분리한 DNA 단편에 카로티노이드 생합성에 필요한 단백질을 코딩하는 *crt* 유전자가 존재함을 확인할 수 있었다(도 8 참조).

<70> 본 발명의 또 다른 실시예에서는 본 발명에서 분리된 *crt* 유전자를 카로티노이드를 생산하지 않는 대장균에 도입하여 상기 *crt* 유전자에 의해 암호화되는 각 단백질에 의해 대장균에서 카로티노이드가 생산되는지 확인하였다. 이를 위해, *crt* 유전자가 삽입된 재조합 벡터 pCR-XL-TOPO-crtfull을 제작하였다(도 9 참조). 이후, 상기 재조합 벡터를 대장균에 도입하여 오렌지색을 띠는 형질전환체를 선발하였다. 상기 형질전환체가 아스타잔틴을 생산하는지 확인하기 위하여, 상기 형질전환체를 배양하여 균체를 회수한 후, 메탄올을 첨가하여 4℃에서 하룻밤 동안 배양하였다. 이후, 상동액을 수득하여 190-900nm에서 흡광도를 측정하였다. 그 결과, 도 10에 도시된 바와 같이,  $\beta$ -카로텐과 아스타잔틴의 고유 피크를 확인할 수 있었다. 보다 정확한 분석을 위하여, 상기 상동액의 일부를 취하여 HPLC 분석을 수행하였다. 이 때 표준물질로는 시그마(Sigma)에서 구입한  $\beta$ -카로텐과 아스타잔틴을 사용하였다. 그 결과, 본 발명에 따라 분리된 *crt* 유전자가 도입된 형질전환체가  $\beta$ -카로텐과 아스타잔틴을 생산함을 확인할 수 있었다. 본 발명의 형질전환체에서 생산된 아스타잔틴의 양은 110  $\mu\text{g/g}$ (dry weight)이었다. 이는 아스타잔틴을 생산하는 균주인 파라코커스 해운대시스가 아스타잔틴을 25  $\mu\text{g/g}$ (dry weight) 정도 생산하는 것과 비교할때에 *crt* 유전자를 대장균에 도입함으로써 아스타잔틴을 훨씬 더 대량으로 생산할 수 있음을 나타내는 결과이다.

<71> 이하, 본 발명을 실시예에 의해 상세히 설명한다.

<72> 단, 하기 실시예는 본 발명을 예시하는 것일 뿐, 본 발명의 내용이 하기 실시예에 한정 되는 것은 아니다.

<73> <실시예 1> 카로티노이드 생합성 유전자의 클로닝을 위한 게노믹 DNA의 준비

<74> 파라코커스 해운대시스(KCCM-10460)을 PPES-II 배지(트립تون 1 g/ℓ, 박토-소이톤 1 g/ℓ, 폐릭 시트레이트 0.01 g/ℓ, 폴리펩톤 2 g/ℓ 및 염화나트륨 3 g/ℓ)에서 25°C에서 10일 간 배양하였다. 이후, 배양액을 13,000 rpm에서 원심분리하여 세포를 수득하였다. 이후, 세포로부터 게노믹 DNA를 분리하기 위해, STE 완충용액(10 mM Tris, 1 mM EDTA, 100 mM NaCl, pH 8.0)에서 세포를 혼탁시킨 후, 68°C에서 15분간 반응시켰다. 원심분리하여 세포를 수득한 후, 용액 I(50 mM 글루코스, 25 mM Tris, 10 mM EDTA, pH 8.0)에 재현탁시켰다. 5 mg/ml 라이소자임(lysozyme)과 100 µg/ml RNase A를 첨가하고 1시간 동안 37°C에서 반응시켰다. 그리고 나서, 단백질 분해효소(Proteinase K)를 250 µg/ml 되도록 첨가한 후, 37°C에서 3시간 동안 다시 반응시켰다. N-라우로일살코신(N-lauroylsarcosine)을 총부피의 1%가 되도록 넣고, 37°C에서 반응시켰다. 이후, 폐놀-클로로포름 추출법(phenol-chloroform extraction)에 따라 게노믹 DNA를 분리하였다. 동일 부피의 폐놀-클로로포름을 넣어 추출한 뒤 2배 부피의 100% 에탄올을 넣어 게노믹 DNA를 침전시킨 후, 70% 에탄올로 세척하였다. 그리고 나서, TE 완충용액을 첨가하여 65°C에서 완전히 녹인 후, 이후의 실험에 사용하였다.

<75> <실시예 2> 게노믹 DNA 라이브러리(Genomic DNA library) 구축

<76> <2-1> 코스미드 벡터(cosmid vector) 준비

<77> 코스미드 벡터(SuperCos 1 Cosmid Vector, Stratagene) 25  $\mu\text{g}$ 에 *Xba* I(9 U/ $\mu\text{g}$ ) 제한효소를 첨가하고, 총 반응액이 200  $\mu\text{l}$ 가 되게 하여 37°C에서 1시간 동안 반응시켜 절단하였다. 이후, 폐놀-클로로포름 추출법에 따라 벡터 DNA를 분리한 후, 100% 에탄올로 침전시켰다. *Xba* I으로 절단된 벡터를 탈 인산화(dephosphorylation)시키기 위해 CIAP 효소(Promega)를 첨가하여 37°C에서 30분 동안 반응시켰다. 이후, 다시 폐놀-클로로포름 추출법으로 벡터를 분리한 후, 100% 에탄올로 침전시켰다. 상기 분리된 벡터를 다시 *Bam*H I(5 U/ $\mu\text{g}$ )으로 37°C에서 1시간 반응시킨 후, 폐놀-클로로포름 추출법으로 분리하고 에탄올로 침전시켰다. 이후, TE 완충용액에 용해시켜 1  $\mu\text{g}/\mu\text{l}$ 가 되도록 하였다.

<78> <2-2> 게노믹 DNA 라이브러리 구축

<79> 상기 실시예 1에서 얻은 파라코커스 해운대시스의 게노믹 DNA 100  $\mu\text{g}$ 에 *Sau*3A I(10 U)을 처리하여 부분적으로 효소 반응시켰다. 반응 종결 후, 0.5 M EDTA를 첨가하였다. 이후, 폐놀-클로로포름 추출법으로 게노믹 DNA를 분리한 후, 100% 에탄올로 침전시켰다. 부분 효소 반응된 게노믹 DNA를 TE 완충용액에 용해시키고, 상기 실시예 1과 동일한 방법으로 CIAP 효소를 처리하여 탈 인산화시켰다. 다시 폐놀-클로로포름 추출법으로 DNA를 분리한 후, 상기 실시예 <2-1>에서 준비된 코스미드 벡터와 라이게이션(ligation)시키기 위해, T4 라이게이즈(ligase, Promega), 10 >라이게이즈 완충용액(Promega)을 넣고 12°C에서 18시간 동안 반응시켰다. 반응 후 상기 라이게이션 혼합액(ligation mixture)을 대장균인 XL1-Blue(Stratagene)에 형질도입하여 게노믹 DNA 라이브러리를 제작하였다.

<80> <실시예 3> 색소 생성 유전자를 함유하는 형질도입 균주의 탐색 및 분석

<81> LB agr 배지에 카로티노이드 계통의 공통된 기질 중의 하나인 FPP(Sigma)를 1%가 되도록 첨가한 후, 상기 실시예 2에서 제조된 게노믹 라이브러리를 상기 플레이트에 도말하여 37℃에서 배양하였다. 배양된 콜로니 중에서 오렌지색을 띠는 13개의 콜로니를 선별하였다(약 2000 개 중 13개). 상기 선별된 13개의 콜로니로부터 코스미드 벡터를 분리하였다. 이후, 프라이머 워킹 시퀀싱(primer working sequencing)을 수행하여 각 코스미드 벡터에 삽입된 DNA 단편의 염기서열을 결정하였다. 이 때 염기서열의 결정은 (주)제노텍에 의뢰하여 수행하였다.

<82> 그 결과, 코스미드 벡터에 삽입된 DNA 단편 중에서 가장 작은 크기의 DNA 단편은 6,223 bp이었으며, 서열번호 1로 기재되는 염기서열을 가졌다. 이에 상기 서열번호 1로 기재되는 염기서열을 단편으로 포함하고 있는 코스미드 벡터를 'COSCRT'라 명명하였다.

<83> <실시예 4> 카로티노이드 생합성 관련 유전자를 포함하는 DNA 단편의 서열 분석

<84> 상기 실시예 3에서 얻은 DNA 단편의 서열을 NCBI ORF Finder 프로그램 (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>)을 사용하여 ORF를 분석하였다.

<85> 그 결과, 상기 실시예 3에서 얻은 DNA 단편에는 6개의 ORF가 포함되어 있었다. 각 ORF는  $\beta$ -카로텐 케토레이즈( $\beta$ -carotene ketolase)를 코딩하는 *crtW*,  $\beta$ -카로텐 하이드록실레이즈( $\beta$ -carotene hydroxylase)를 코딩하는 *crtZ*, 리코펜 싸이클레이즈(lycopene cyclase)를 코딩하는 *crtY*, 피토엔 디세튜레이즈(phytoenedesaturase)를 코딩하는 *crtI*, 피토엔 신테이즈(phytoene synthase)를 코딩하는 *crtB* 및 제라닐제라닐 피로포스페이트 신테이즈

(geranylgeranyl pyrophosphate synthase)를 코딩하는 *crtE*의 염기서열과 높은 상동성을 나타내었다. 상기 효소들은 모두 카로티노이드 생합성에 관여하는 효소들이다.

<86> 또한, 각 ORF로부터 유추되는 아미노산 서열과 알칼리제네스 속(*Alcaligenes* sp.)(Misawa *et al.*, *Biochem. Biophys. Res. Commun.*, 209(3): 867-876, 1995), 브라디리조비움 속(*Bradyrhizobium* sp.)(Hannibal *et al.*, *J. Bacteriol.*, 182(13): 3850-3853, 2000) 및 플라보박테리움 속(*Flavobacterium* sp.)(Pasamotes *et al.*, *Gene*, 185(1): 35-41, 1997)에서 분리된 각 카로티노이드 생합성 효소들의 아미노산 서열과의 상동성 비교 결과를 도 2 내지 도 7에 나타내었다. 이에, 본 발명에서 클로닝된 6개 ORF 유전자의 염기서열 및 이로부터 유추되는 아미노산 서열을 서열번호 2 내지 13으로 각각 기재하였다. 즉, *crtW*, *crtZ*, *crtY*, *crtI* 및 *crtB*의 유전자는 각각 서열번호 2, 서열번호 4, 서열번호 6, 서열번호 8, 서열번호 10 및 서열번호 12이며, 상기 유전자 각각의 아미노산 서열은 서열번호 3, 서열번호 5, 서열번호 7, 서열번호 9, 서열번호 11 및 서열번호 13이다.

<87> 상기 결과로부터 상기 코스미드 벡터에 삽입된 DNA 단편에 카로티노이드 생합성에 관여하는 *crt* 유전자가 존재함을 확인할 수 있었다.

<88> 본 발명의 *crt* 유전자의 구성을 도 8에 도시하였다. 도 8에 도시된 바와 같이, 각 *crtW*, *crtZ*, *crtY*, *crtI* 및 *crtB*는 종결 코돈과 시작 코돈이 중첩되고 있음을 볼 수 있다. 특히, *crtE* 유전자는 상보적 가닥의 방향성을 가지는 것을 볼 수 있으며, 서열 내에 *Kpn*I, *Xma*I, *Sma*I, *Cla*I, *Hind*III 및 *Bam*H I 인식 서열이 하나씩 존재함을 볼 수 있다.

<89> <실시예 5> 대장균에서의 *crt* 유전자의 발현 실험

<90> 본 발명자들은 상기 실시예 3에서 분리한 파라코커스 해운대시스의 *crt* 유전자로부터 발현된 단백질들에 의해 카로티노이드가 생성되는지 확인하기 위하여, 먼저, 상기 *crt* 유전자를 HL프리믹스(HLpremix, Bioneer)를 이용하여 PCR로 증폭하였다. 이 때 프라이머로는 서열번호 14 및 15로 기재되는 올리고뉴클레오티드를 사용하였다. 또한, PCR은 94°C에서 5분 동안 주형 DNA를 전변성화시킨 후, 68°C에서 1분, 72°C에서 6분 한 후 94°C에서 30초, 66°C에서 30초 및 72°C에서 6분을 한 사이클(cycle)로 하여 총 25 회 반복수행한 다음, 마지막으로 72°C에서 20분 동안 반응시켜 수행하였다. 이후, PCR 산물을 Topo-XL-vector(invitrogen)에 삽입한 후, 이를 대장균에 형질도입하였다. 이 때 대장균으로는 XL1-Blue(Stratagene), TOP0(Invitrogen), BL21(DE3) 코돈 플러스(Stratagene), DH1(Takara) 및 DH5 *a* (Takara)를 사용하였다. 그 결과, BL21(DE3), XL1-Blue, BL21(DE3) 코돈 플러스 형질전환체들이 오렌지색을 띠는 것을 확인할 수 있었다. 이후, 각 대장균에 대하여 형질전환체 하나씩을 선별하여 배양한 결과, 형질전환된 BL21(DE3) 코돈 플러스가 가장 많은 아스타잔틴을 생산해 내는 것을 확인할 수 있었다.

<91> 이후, 본 발명자들은 상기 실시예 3에서 분리한 서열번호 1로 기재되는 염기서열을 가지는 *crt* 유전자를 유전자 발현용 벡터인 pCR-XL-TOP0 벡터(Invitrogen)에 삽입하여 이를 'pCR-XL-TOP0-crtfull'이라 명명하였다(도 9). 이후, 상기 pCR-XL-TOP0-crtfull 벡터를 BL21(DE3) 코돈 플러스 세포에 형질도입시킨 후, 오렌지색을 띠는 균주를 선별하였다. 선별된 균주를 50 ml LB 배지에서 37°C, 8시간 동안 배양하였다. 이후, 배양액을 13,000 rpm에서 원심분리하여 상등액을 버리고, 균체를 회수하였다. 회수된 균체에 메탄올 20 ml를 첨가하여 볼텍싱(vortexing)한 후, 4°C에서 하룻밤 동안 배양하였다. 이후, 13,000 rpm으로 원심분리한 후, 상등액을 수득하였다. 카로테노이드의 생성 유무를 확인하기 위하여, 190-900 nm의 파장

및 400-550 nm의 파장으로 상기 상등액의 흡광도를 스캐닝하였다. 그 결과, 450 nm와 470 nm에서 피크(peak)를 확인하였으며, 이것이  $\beta$ -카로텐과 아스타잔틴의 고유 피크임을 확인하였다 (도 10).

<92> 보다 정확한 분석을 위해, 상기 상등액 1 mL를 취하여 직경 0.45  $\mu\text{m}$ 의 필터로 여과한 후, HPLC 분석(column: 4.6×50mm, uBondapak C18, Waters, Milford, MA; mobile phase: acetonitrile-methanol-water(49:44:7 v/v), Flow: 10mL/min, Detector: 470nm)을 수행하였다. 이 때 표준물질로는 시그마(Sigma)에서 구입한  $\beta$ -카로텐과 아스타잔틴을 사용하였다. 그 결과, 본 발명의 균주로부터 생산된 물질이 아스타잔틴과  $\beta$ -카로텐임을 확인할 수 있었다.

<93> 또한, 본 발명의 균주로부터 생산된 아스타잔틴의 양을 측정한 결과, 110  $\mu\text{g/g}$ (dry weight)으로 확인되었다. 이상의 결과는 어떤 유도제나 카로티노이드 기질을 첨가하지 않은 상태에서의 결과로서 본 발명에서 분리한 6,223 bp의 염기서열만으로 대장균에서  $\beta$ -카로텐과 아스타잔틴을 대량으로 생산할 수 있음을 알 수 있다. 이는 아스타잔틴을 생산하는 파라코커스 해운대시스(기탁번호:KCCM-10460)가 25  $\mu\text{g/g}$ 의 아스타잔틴을 생산하는 것에 비해 훨씬 많은 양임을 알 수 있다.

### 【발명의 효과】

<94> 상기에서 살펴본 바와 같이, 본 발명에서는 아스타잔틴을 생산하는 파라코커스 해운대시스로부터 카로티노이드 생합성에 필요한 단백질을 코딩하는 6개의 유전자 및 이를 포함하는

*crt* 유전자를 클로닝하였다. 또한, 상기 *crt* 유전자를 이용하여 카로티노이드를 생산하지 않는 대장균에서 카로티노이드를 생산할 수 있음을 규명하였다. 본 발명의 유전자 및 이를 포함하는 *crt* 유전자는  $\beta$ -카로텐, 아스타잔틴과 같은 카로티노이드의 생산에 매우 유용하게 사용될 수 있다.

## 【특허청구범위】

## 【청구항 1】

카로티노이드 생합성에 필요한 단백질을 코딩하는 서열번호 2, 서열번호 4, 서열번호 6, 서열번호 8, 서열번호 10 및 서열번호 12로 기재되는 염기서열로 구성된 군으로부터 선택되는 것을 특징으로 하는 유전자.

## 【청구항 2】

제 1항에 있어서,  $\beta$ -카로텐 케토레이즈( $\beta$ -carotene ketolase)를 코딩하는 *crtW* 유전자의 염기서열인 것을 특징으로 하는 서열번호 2로 기재되는 유전자.

## 【청구항 3】

제 1항에 있어서,  $\beta$ -카로텐 하이드록실레이즈( $\beta$ -carotene hydroxylase)를 코딩하는 *crtZ* 유전자의 염기서열인 것을 특징으로 하는 서열번호 4로 기재되는 유전자.

## 【청구항 4】

제 1항에 있어서, 리코펜 사이클레이즈(lycopene cyclase)를 코딩하는 *crtY* 유전자의 염기서열인 것을 특징으로 하는 서열번호 6으로 기재되는 유전자.

## 【청구항 5】

제 1항에 있어서, 피토엔 디세튜레이즈(phytoene desaturase)를 코딩하는 *crtI* 유전자의 염기서열인 것을 특징으로 하는 서열번호 8로 기재되는 유전자.

## 【청구항 6】

제 1항에 있어서, 피토엔 신테이즈(phytoene synthase)를 코딩하는 *crtB* 유전자의 염기서열인 것을 특징으로 하는 서열번호 10으로 기재되는 유전자.

## 【청구항 7】

제 1항에 있어서, 제라닐제라닐 피로포스페이트 신테이즈(geranylgeranyl pyrophosphate synthase)를 코딩하는 *crtE* 유전자의 염기서열인 것을 특징으로 하는 서열번호 12로 기재되는 유전자.

## 【청구항 8】

제 2항 내지 제 7항의 유전자를 모두 포함하는 서열번호 1로 기재되는 염기서열을 갖는 것을 특징으로 하는 *crt* 유전자.

## 【청구항 9】

제 8항의 *crt* 유전자를 포함하는 재조합 벡터.

## 【청구항 10】

제 9항에 있어서, 도 9에 도시된 개열지도를 갖는 것을 특징으로 하는  
pCR-XL-TOP0-crtfull 재조합 벡터.

## 【청구항 11】

제 10항의 재조합 벡터를 형질도입한 대장균 형질전환체.

## 【청구항 12】

- 1) 제 8항의 *crt* 유전자를 클로닝하는 단계;
- 2) 단계 1의 *crt* 유전자를 포함하는 재조합 벡터를 제조하는 단계;
- 3) 단계 2의 재조합 벡터를 숙주세포에 형질도입하는 단계;
- 4) 단계 3의 형질도입된 균주를 배양하여 그 배양액으로부터 카로티노이드를 회수하는 단계인 것을 특징으로 하는 카로티노이드를 생산하는 방법.

## 【청구항 13】

제 12항에 있어서, 상기 재조합 벡터는 제 10항의 재조합 벡터인 것을 특징으로 하는 방법.

## 【청구항 14】

제 12항에 있어서, 상기 숙주세포는 대장균 또는 효모인 것을 특징으로 하는 방법.

## 【청구항 15】

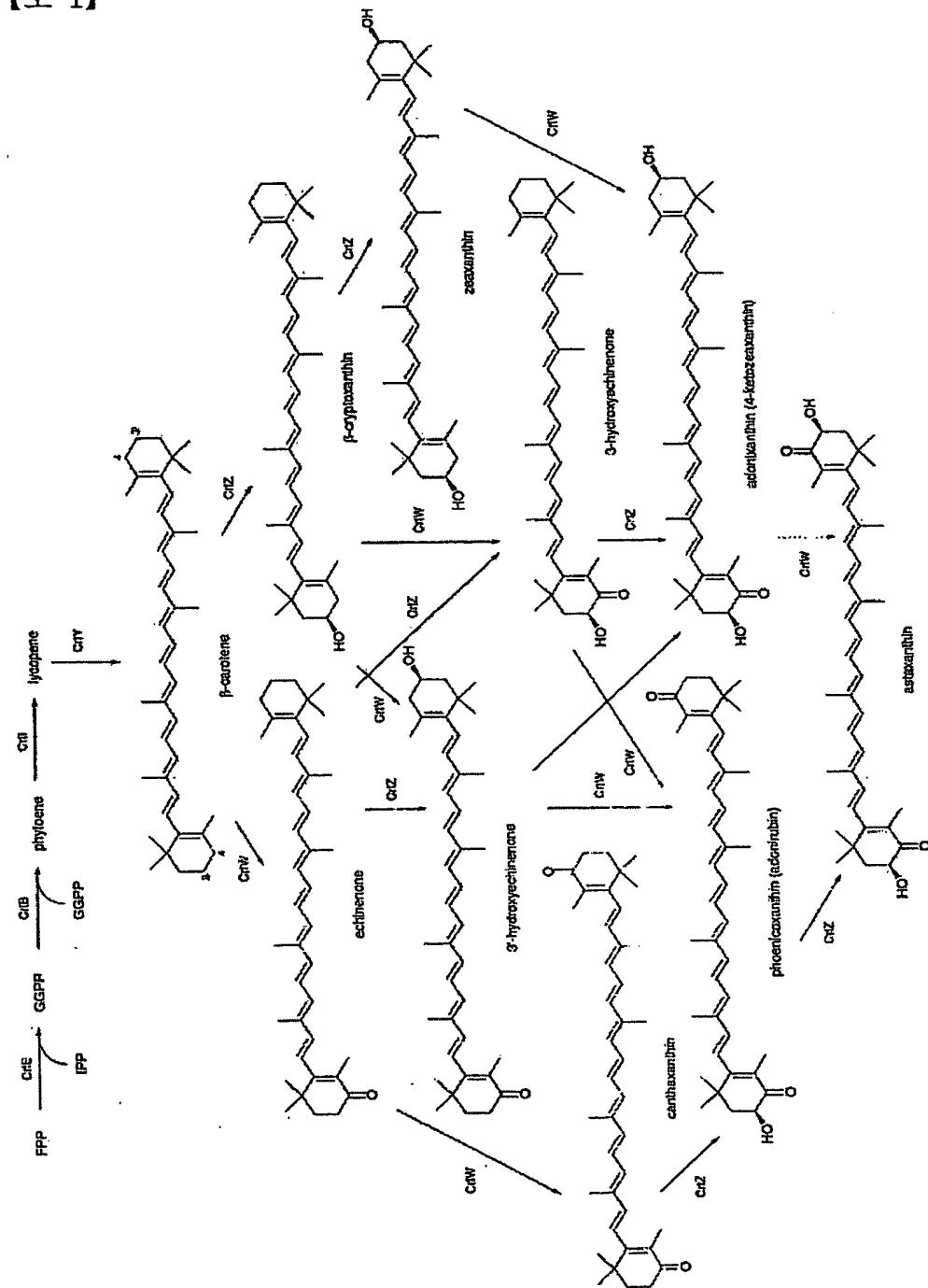
제 12항에 있어서, 제 11항의 대장균을 배양하여 그 배양액으로부터 카로티노이드를 회수하는 것을 특징으로 하는 방법.

## 【청구항 16】

제 12항에 있어서, 상기 카로티노이드는  $\beta$ -카로텐 또는 아스타잔틴인 것을 특징으로 하는 방법.

### 【도면】

【도 1】



## 【도 2】

1 100

P. haemadensis (1) -----NSABALPKADLTATSLIVSGGIIAAALALBVBALVPLDAAAAPILAIANFLGLT-VLSVGLFFIAADAMHGJVVPGPDRGHAANGOLVIV  
 Alcaligenes\_sp (1) -----NSGKPGTGTGIVIIVNLGLTAAKLLCBLVBLAFTLULLDAAAPBLAVLCLAGLT-VLSVGLFFIAADAMHGJVVIGPDRBAAAIQOLALV  
 Bradyrhizobium\_sp (1) NHAAATAKATEFGASREDDARQRRVGLTLAAVIIAAVLVLVHGLNFFVPLTLBSLLPAPLPLVVLQTLVYVGLFTIAADAMHGSLVPPFKPQVNRRIQQLCLF  
 Consensus (1) NSAAKATIVLSSAIIIAAVLVLVLYLWFLLAAAPLLAILLNLGLT-VLSVGLFFIAADAMHGJVVPGPDRBAAAIQOLLV

101 200

P. haemadensis (90) LYAGFSVRLVHIVKHNHRRHNTGDDDPDFDGG--GPVRVYARFIGTYFGVREGLLLPVIIVVYALILGD-BWVYVVFVPLPSILASIQLFVFGTIVLPHE  
 Alcaligenes\_sp (90) LYAGFSWPKLIAKHTHDSBAGTDDDPDFDGG--GPVRVYGSFVSTYFGVREGLLLPVIIVTVALILGD-BWVYVIFVDFVPLASIQLFVFGTIVLPHE  
 Bradyrhizobium\_sp (101) LYAGFSFDALVYEHKHNHRSPTAEDIDPDEVPPHGFVHNPASFLBLHFGKQVALLAAASVLYVQVFAVPLQNLILPFLVPLGILSALQLFTOTVLPHE  
 Consensus (101) LYAGFSVHLIVHIVHRRBGTDDDPDFDGGGPVBUYASFIYFGVREGLLLPVIIVVYALILGD-BWVYVIFVPLPSILASIQLFVFGTIVLPHE

201 239

P. haemadensis (186) PGHDAPPDRHMASRSISDPVSLLTICFHFGGYHREHHLHPTVPLPSTRTKGDTA--  
 Alcaligenes\_sp (186) PGHDAPPDRHMASRSISDPVSLLTICFHFGGYHREHHLHPTVPLPSTRTKGDTA--  
 Bradyrhizobium\_sp (201) PATQPFADRHMASRSISDPVSLLTICFHFGGYHREHHLHPTVPLPSTRTKGDTA--  
 Consensus (201) PGHDAPPDRHMASRSISDPVSLLTICFHFGGYHREHHLHPTVPLPSTRTKGDTA

## 【도 3】

1 100

P. haemadensis (1) HTNFLIVVATVLYNELTAYSVHRYVIMHQLQGVQVHKSHEEDBALKNDLYGLVPAVIAITVLFIVGVIVAPVLYQVIALGHTVGLIYFVLHIAJLVHQEW  
 Alcaligenes\_sp (1) HTQFLIVVATVLYNELTAYSVHRYVIMHQLQGVQVHKSHEEDBALKNDLYGVWVPAVIAITVLFIVGVIVAPVLYQVIALGHTVGLIYFILHIAJLVHQEW  
 Consensus (1) HTNFLIVVATVLYNELTAYSVHRYVIMHQLQGVQVHKSHEEDBALKNDLYGLVPAVIAITVLFIVGVIVAPVLYQVIALGHTVGLIYFILBDGLVHQEW

101 163

P. haemadensis (101) PFBYIIPREGYARBLYQAHRLBHAVEGRDHCYCSFGFIYAPPVDELKQDLKTSGVLAEEQEBT  
 Alcaligenes\_sp (101) PFBYIIPREGYFBRBLYQAHRLBHAVEGRDHCYCSFGFIYAPPVDELKQDLKTSGVLRPQDERPS  
 Consensus (101) PFBYIIPRKGYBRLYQAHRLBHAVEGRDHCYCSFGFIYAPPVDELKQDLKTSGVLRPS

## 【도 4】

1 100

P. haemadensis (1) VHDVLLAGACGLANGLIALALBAPDLVPLLLDHAAGPSDGTWSCHDPDLSPHULARLPLBANLPGDQEVRPPRHARLNLYGCELDGAAALADAVAR  
 Flavobacterium\_sp (1) NSHDLLIAGAGLALQSLALAVEDBRRPAKLVVLDPARSOPSDQHWSCHDILSPFELALSPIRGEVTDQEVAFPDSRRLTQGSIEGALIGLLO-  
 Consensus (1) NSHDLLIAGAGLALQSLALAVEDBRRPAKLVVLDPARSOPSDQHWSCHDILSPFELALSPIRGEVTDQEVAFPDSRRLTQGSIEGALIGLLO-

101 200

P. haemadensis (101) SGAEIRHNSDIALDEQJAYLSCGTYRIBAGAVYLDGRGQPSRHLTVGPQKFVGVETIETLCPHGVPRPHINDATVQQDQYRFTIVYLLPF3FTBILIEDTBY  
 Flavobacterium\_sp (100) -GVDLRENTBVATLDDTGATLTDGSRIEAAACVYDARGAVETPHLTVGFQKFVGVETIETDAPRGVERPHINDATVQMDGYRFTIVYLLPFSPTRBLIEDTBY  
 Consensus (101) GDIRVNSIA LDG GATLS GSRIEAA VIDARGA S HLTVGQKFVGVETIETD PHGV BPMHINDATVQDGYRFTIVYLLPFSPTRBLIEDTBY

201 300

P. haemadensis (201) SDGQHLDODALAASHDYARQQGFTGAEVBRERBGILVTLALDAA3FVADRAEGPVYVGLRACEFPRPVYQTSIPLVYAOVADVYAGLSGPPTGDAEAGAK  
 Flavobacterium\_sp (199) SDGQHLDODGALAAQSLDYAARRGUTQBRERGERTGILPITALDAAIGFURDHAQGAVPVULGACLFHIVIGYSLPYAAQVADAAAR--DLTASAKRAVE  
 Consensus (201) SDGG LDD ALA AS DVA GFTG ERREERBGILPIALABDA GFT DHA G VPVGL AG FBPVIGYSLPYAAQVAD IAA T A R AIR

301 387

P. haemadensis (301) DYAIDARRDKPFLRLLNRHLFSGCAPDDEBYTLLQRFYEMPBGLIERFYAGRLSVADQLRIVIGKPPFLGTAIRCLPERPLLKEA  
 Flavobacterium\_sp (297) GWAIPSDADRLRPLRLLNRHLFSGCPPDBKYLQKFTKLQPPLIERFYAGRLTLADRLTIVIGRPPPLSQAVRCLPERPLLQSGA  
 Consensus (301) DAIIDRA BDRFLRLLNRHLFSGC PDRRY LLAQRFYRLP LIERFYAGRLSLAD LRIVIGKPPPL AIRCLPERPLL

## 【도 5】

1 100

P. haeundaeensis (1) WSAHSPAAKTYIYI GACFGGLALAIRLQSGAI ATTIVAEARDPEGGRAVYVHDGHVFDAGPTVITDPDAKELVALTGQDHARDVILMPVSPPFYRLWAD

Flavobacterium\_sp (1) -----HSSAIVIGAAGFGGLALAIRLQSGAI ATTIVAEARDPEGGRAVYVHDGHVFDAGPTVITDPDAKELVALSGGPHEDVILPPVSPPFYRLWAD

Consensus (1) S IVIGAAGFGGLALAIRLQSGAI ATTIVAEARDPEGGRAVYVHDGHVFDAGPTVITDPDAKELVALSGQ W RDVILPPVSPPFYRLWAD

101 200

P. haeundaeensis (101) GKVFVDYVHEADQLEBQIAQFNPPDLEGYERFRDVAEEVYQEGYVVLGTVPFLKLQHLLKAPALHLEAVESVHAKVATIFIKDPYLBQAFSYTLLVGD

Flavobacterium\_sp (94) GRSFEVYVNDDEDFLREQVASEFNPADYDQYRFRHDVAEEVYREGYVLLGTIPFLKLQHLLKAPALHLEAVESVHNSYARPIQDPHLBQAFSPYTLVGGH

Consensus (101) GK FDYVND D L BQIA FNP DLDGYRBP DYAEVY EGVLKLGT PFLKLQML AAPALMEL AVESVHA VA FI DPHLBQAFSPYTLVGGH

201 300

P. haeundaeensis (201) PFSTSSIVALIHALERBRGGVVFAGQNTNQLVAGHVALPERLGGQHHLBKVAKLETPGARTTGYVTLADGSSLRADVVASBGDVHETVPLAHTARGQSB

Flavobacterium\_sp (194) PFSTEDIVALIHALERBRGGVVFAGGTNQLVAGHVALPERLGGTLLLNHARVTEIDIEGDRATVGYVTLDEQRLRADTVASBGDVHESYRDLHETREGRTK

Consensus (201) PFSTSSIVALIHALERBRGGVVFAGQNTNQLVAGHVALPERLGG TLLNKA VIEDEG R TGYVTL DGR L EAD VASNGDVHESYRDLHETRGRTK

301 400

P. haeundaeensis (301) AKSLDRKKWMSVFLVLMFQLREAPKDIABNTILEFGPRYRBLVSELPFGPKLAEDFSLYLBSPCTTDPDMPGQHSTHYVLAPVPLIKAEDIVWAEQGPY

Flavobacterium\_sp (294) AAILNBQRWSMSLFLVLMFQLSKRPNLADNSVIFGPRYKGLVNETPNGRLPDDPSVYLBSPCVTDPSLAEGHSTHYVLAPVPLGKADWDVVAEAPGY

Consensus (301) A L B RWSMSLFLVLMFQL P IAHESIIFGPRYK LVNEIF GPKL DDESLYLBSPC TDP LAP GHSTHYVLAPVPLGKADWDVVAEAPGY

401 500

P. haeundaeensis (401) ADRLLASLEERLIPNLPANLITTRIFTPAPDEASBLNAHGSAFSVEPLITQGAVPRPHHRDTIIRGFLVAGCTHFGAGIPGVVVGSAKATAQVHLSLAG

Flavobacterium\_sp (394) AEKI FEELERBAIPDLKHLTVSKIFSPADESTELSAHGSAFSVEPLITQGAVPRPHEADRAIIPNFYIVGAGTHFGAGIPGVVVGSAKATAQVHLSLAG

Consensus (401) ADRI LE B IP LR LT SRIFSPADFASEL AHGSAFSVEPLIQS AVEPRPHNEDK I NFYIVGAGTHFGAGIPGVVVGSAKATAQVHLSLAG

501

P. haeundaeensis (501) A

## 【도 6】

1 100

P. haeundaeensis (1) WSDLVLTSEATIQQQSFAIAKELHPGIRDPTVHLVACBEGADDVTDGQALGSRPEAVNDPQARCLDGLRVDTLAALQDGPVTPPEAALBAVABBDHF

Flavobacterium\_sp (1) HYLDTATEEAIWQSFAQAAKLMPPQIREDITVHLVACBHDVILGQVNGSAPENAGDPQAEGLALBDTIAALHEDGPHSVPFPAAALRQVABBDHF

Consensus (1) WSDL TS AI QGSQSFAAAKLMPPQIREDITVHLVACBHDVIDQ LGS PEA DPQARL ALB DTIAAL DGPMSPPFAALB VARBDP

101 200

P. haeundaeensis (101) PQASITWILIEGFANDVBEARDTSTLDDVLEYCYHNGIVGVHMAEVNGVRODPLDRACNLGLAFQLTIAIDVTDARIGBCYLFGDPLDQAGARIDGPV

Flavobacterium\_sp (101) PDLPPHLLIEGFANDVADBEYRSLLDVLEYTSYHNGIVGVHMAEVNGVQNDVYLDRACNLGLAFQLTIAIDVTDAAIGBCYLPADVLAEGATVEGPV

Consensus (101) P UPNDLIEGFANDV RQYRSLLDDVLEYTSYHNGIVGVHMAEVNGV DD VLDACDLGLAFQLTIAIDVTDAAIGBCYLPADVLAEGATVEGPV

201 300

P. haeundaeensis (201) ISPELVTVILRLLDEAEPYVSAKRGADLPPCANSIAAALRIYRAIGLRIEKSQGPQAYEQKIIISKAATIGELGVGGDVARSKLPQAGVSGQSLWTR

Flavobacterium\_sp (201) FEDALYSVIIILRLLDEAEPYVSAKQGLPHELPPCANSIAAALRIYRAIGLRIEQQGPPEAYQKRISTSKAATIGELLABQGLDAASSLRQSGEICRQDILWTR

Consensus (201) PS LVSVIILRLLDEAEPYVASAR GL LPPECANSIAAALRIYRAIG RIK GP AYQKRISTSKAATIGELLA GG D A SEL GA ISH GLWTR

301

P. haeundaeensis (301) PBBV

Flavobacterium\_sp (301) PRA-

Consensus (301) P

## 【도 7】

1 100

P. haeundaeensis (1) HERDVPYTHAIIQLTALIEEIAQFGACVQPLGAAHSBGMALSGRSRFRGHLHLLAAEASGGVCDTIVDAAACAVENVHAASLI FDDLPCHDDAQLKRGPRAT

Flavobacterium\_sp (1) HTPKQQFPRLDLYEIRLAQISGQPGVYVAPLGAMMCSDALCPQKREFRAVLMLNVAESGCGWCDAMVAAACAVENVHAASLIFDQFCCHDABTRRQCPAT

Consensus (1) H R LL BL IA FG VS PLAGAAS AALS GKRFRAMLLA AEASGGVCD IVDAACAVENVHAASLIFDLPCHDDA RRG PAT

101 200

P. haeundaeensis (101) BVAHGEKAVLGKIALITEAHALLAGAEGGAGSTVKAQLVRLIBSLIYQGLCAQQLDQHAAKRGAGQVEEQDLKTVLFIAGLFHLAVIKEFDAEQTQ

Flavobacterium\_sp (101) BVAHGEGRVAVLAIALITEAHLLAGAEGGAGTDPQKARLVAHSNSKAMGPVQICLCAQQLDQHAAKRGAGQVEEQDLKTVLFIAGLFHLAVIKEFDAEQTQ

Consensus (101) BVAHGE BAVLGKIALITEA ILA ARGAS RA LV LSRALGP GLCAQQLDQHAAKRGAGQVEEQDLKTVLFIAGLFHLAVIKEFDAEQTQ

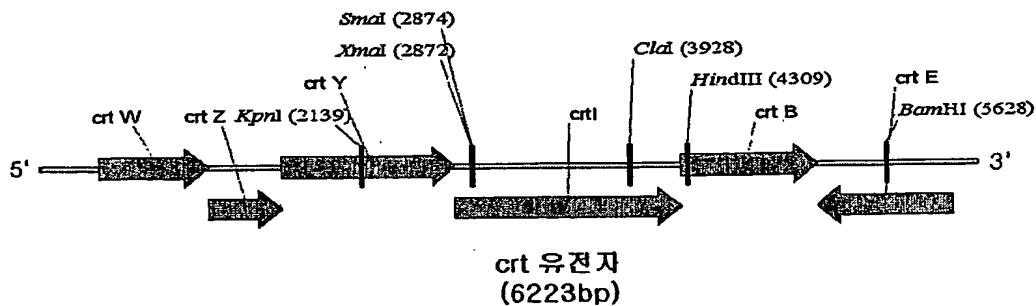
201 295

P. haeundaeensis (201) MIFDQICOLGRVFGQYDNLIDVYVQDQAAQGKDTGCDAAAPGDPBRGQLLAVSDIQLNYSRHYEASEMOLDAMLRSLRQPEIALLERVLVLPYAA--

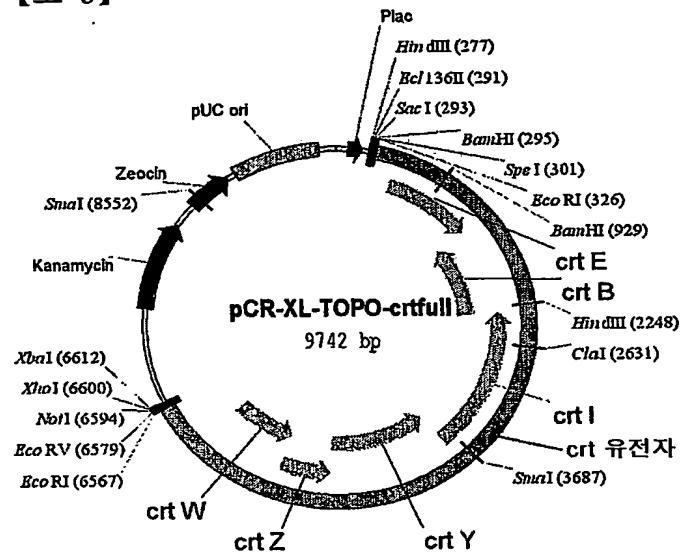
Flavobacterium\_sp (201) LNAGURQLCGVFGQYDNLIDVYVQDQKASTGKUTARLTAPGSPKQELNVAQGQUD7AQHRYRACRQQLDHLHETRLPRGGQIADLLARVLPHDIIKRS

Consensus (201) LI FGRQLGRVFGQYDNLIDVGD AA GKDABD AAPGPK GLLAV L VA BY ASRAGLD LLRSK A IA LL RVLPH B

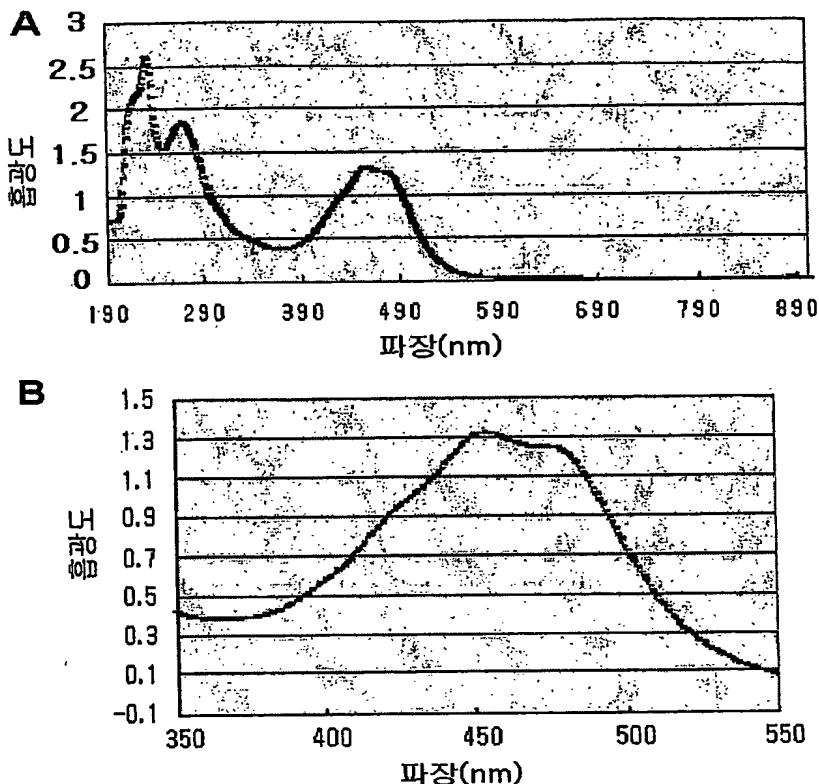
【도 8】



【도 9】



## 【도 10】



## 【서열목록】

<110> KIM, Young Tae LEE, Jae Hyung ALGENETECH <120> Genes involved in the biosynthesis of carotenoids <130> 3p-01-31 <160> 15 <170> KopatentIn 1.71 <210> 1 <211> 6223 <212> DNA <213> crt gene <400> 1 gttccacgac tggggcatcc ccacgaccgc gtcgctgcgc gccatcgccgc  
cgatgtatggg 60 gccggaccgg gttctggtcg ggtcggccgg ggtgcgtcac gggctggacg ccgcgcggc 120  
catccgcctc ggccgcggacc tcgtggggca ggcggcccgcc ggcgtgcgc cccgcgcgcca 180 cagcgccgag gcccgtccg  
atcacctgtc cgacgtcgac acccagctgc gcatcgcat 240 gttctgcacc ggatcggcgc accttgcagc gtcgcgtgc  
gcccctctgc tgggtccggg 300 gcccggatgc caatggtcgc aagcaacggg gatggaaacc ggcgtatgcgg gactgttagtc  
360 tgcgcggatc gcccggatccgg gggacaagat gaggcgcacat gcccgtccca aggcagatct 420 gaccggccacc agcctgatcg  
tctcggccgg catcatcgcc ggcgtggctgg cccatgcgt 480 gcatgcgtg tggttctgg acgcggccggc gcatccatc  
ctggcgatcg cgaatttcct 540 ggggctgacc tggctgtcgg tcggctgtt cttcatcgat catgacgcga tgcacgggtc  
600 ggtcgtgcgg gggcgtccgc gcccgtatgc ggcgtatggc cagctggtcc tggctgtta 660 tgccggattt tcgtggcgcga

agatgtatcgtaaagcacatggcccatcaccgcataccgg 720 aaccgacgac gaccccgatt tcgaccatgg cggcccggtc  
 cgctggtacg cgcgcttcat 780 cggcacctat ttccggctggc gcgagggct gctgctgccc gtcatcgtaa cggtctatgc  
 840 gctgatcctggggatcgct ggtatgtacgt ggtcttctgg ccgctgcccgt cgatcctggc 900 gtcgatccag ctgttcgtgt  
 tcggcacctg gctgccgcac cgcccccggcc acgacgcgtt 960 cccggaccgc cataatgcgc ggtcgctcgac gatcagcgac  
 cccgtcgac tgctgacactg 1020 ctttacattt ggtggttatc atcacaaca ccacctgcac ccgacggcgc ctgggtggc  
 1080 cctgcccacg acccgacca aaaaaaaaaaaaaaaacacacatgacca atttcctgat cgctgctgccc 1140 accgtgctgg tggatggagg  
 gacggcctat tccgtccacc gttggatcat gcacggcccc 1200 ctgggctggg gctggcacaa gtcaccac gaggaacacg  
 accacgcgtt ggaaaagaac 1260 gacgtgtacg gcctggctt tgccgtgatc gccacggcgc tggatggatc  
 1320 tggcgccgg tcctgtggtg gatcgcttgc ggcacatgaccc tctatggct gatctatttc 1380 gtcctgcattt acgggctgg  
 tcatcagcgcc tggccgttcc gctatatccc ggcacatggggc 1440 tatgcccccc gcctgtatca ggcaccac gtcaccac  
 cggatcgaggacat 1500 tgcgtcagct tcggcttcat ctatgcgcgg ccggatcgacca agctgaagca ggacactgaag  
 1560 acgtcggcgc tgctgcggc cgaggcgcag gagcgcacgt gacccatgac gtgtgtgg 1620 caggggcggg cttgcgaac  
 gggctgatcg ccctggcgct ggcgcggcggc cggccggacc 1680 tgcgggtgt gctgtggat catgcggcgg gaccgtcaga  
 cggccataacc tggcctgccc 1740 acgaccccgatc tctgtcgccg cactggctgg ccggatcgacca gcccctgcgc cgcgcacact  
 1800 ggcacca ggaggtgcgc ttccccccgc atggccggcg gctggccacc ggttacgggt 1860 cgctggacgg ggcggcgt  
 gcggatcgcc tggccggcgc gggccggcggc atccgctgg 1920 acagcgacat cgcctgcgt gatgaacagg gggcgacgc  
 gtccctgcggc acccgatcg 1980 aggccggcgc ggtccgtggac gggccggcgc cgcacccgtc ggcacatcg accgtgggtt  
 2040 tccagaaatt cgtggcgatcg gagatcgaga ccgactgcggcc ccacggcgatcg ccccgcccg 2100 tggatcgatcg  
 acccaggcagg acggatcgatcg attcatctat ctgtgcctt 2160 tctctccgac ggcacatcgatcg atcgaggaca ctgcgttattc  
 cgtggcgatcg aatctggatcg 2220 acgacgcgtt ggcggccggc tccacgcgtt atgccccca gcaaggctgg accggggcccg  
 2280 aggtccggcg cgaacgcgcgc atcccgccca ttgcgtggc ccatgacgcgc gcccggatcg 2340 gggccgatca cgcggagggg  
 cctgttcccg tggactcgatcg cgcggggatcg tttcacccgg 2400 tcacccggatca ttgcgtggc ttcggccgc aggtggcgatcg  
 cgtggcgatcg ggcctgtccg 2460 gggccggccgg caccgacgcgc ctgcggccggc ccatccgcga ttacgcgtt gacccggcgc  
 2520 ggcgtgacccg ctttctgcgc ctgtgtatcgatcg ggtatgtatcg 2580 ggcgtatatac ctcgtgcag  
 cgttctacc gcatgcgcgatcg tggactcgatcg gaacggatcg 2640 atgcggccggc gctgagcgatcg ggcgtatcg  
 gcatgcgcgatcg tggactcgatcg gaacggatcg

gaccggcaag cctccattc 2700 cccttggcac ggccatccgc tgcctgccc aacgtccc gctgaaggaa aacgcatgaa  
 2760 cggccattcg cccgccccca agaccgcat cgtgatcgcc gcaggcttg gcgggctggc 2820 cctggccatc cgcctgcagt  
 cgcgggcat cgccaccacc ctggtcgagg cccgggacaa 2880 gccccgggg cgccatcg tctggcacga tcagggccat  
 gtcttcgacg cgggcccac 2940 cgtcatcacc gaccccgatg cgctcaagga gctgtggcg ctgaccggc aggacatggc  
 3000 gcgacgtg acgctgatgc cggtgtcgcc cttcatcgat ctgatgtggc cgggccccaa 3060 ggtcttcgat tacgtgaacg  
 aggccgatca gctggagcgc cagatcgccc agttcaaccc 3120 ggacgacctg gaaggatacc gccgcttcg tgattacgcg  
 gaggaggtgt atcaggaggg 3180 ctacgtcaag ctgggcaccc tgcccttc caagctggc cagatgtca aggccgcgc  
 3240 cgcctgatg aagctggagg cctataagtc cgtccatgcc aaggtcgcga ccttcataa 3300 ggacccctat ctgcggcagg  
 cgttttcgta tcacacgctg ctggtggcg ggaatccctt 3360 ctcgaccagc tcgatctatg cgctgatcca cgcgtggag  
 cggcgcggcg gggctcggtt 3420 cgccaagggc ggcaccaacc agctggtcgc gggcatggc ggcgtttcg aacggcttgg  
 3480 cggccagatg atgctgaacg ccaaggctgc cggatcgag accgagggcg cgcggaccac 3540 gggcgtcacc ctggcggacg  
 ggcggcttt aaggccgac atggtcgcca gcaacggcga 3600 cgtcatgcac aactatcgac acctgtggg ccacacggcc  
 cgcggcaga gccgcgcga 3660 atcgctggac cgcaagcgct ggtccatgtc ttgttgcgt ctgcatttcg gtctgcgcga  
 3720 ggcgcccaag gacatcgccg atcacaccat cctgttcggc cccgcataca gggagctgg 3780 caacgagatc ttcaagggcc  
 cgaagctggc cgaggatttc tcgctgtacc tgcattcgcc 3840 ctgcacgacc gatccggaca tggccctcc gggcatgtcc  
 acgcattacg tgctggcccc 3900 cgtgccat ctggccgcg ccgagatcga ttggcggc gagggccgc gctatgccga  
 3960 ccgcattctg gcgtccctgg aggagccgct gatccgaaac ctgcgcgcca acctgaccac 4020 gacgcgcac ttcacgccc  
 ccgatttcgc cagcgaactg aacgcccac acggcagcgc 4080 ctctcggtc gagccatcc tgacgcaatc cgcgtggttc  
 cggccgcaca accgcacaa 4140 gacgatccgc aacttctatc tggtcggcgc gggcaccat ccggccgcgg gcattccggg  
 4200 cgtcgtggc tcggccaagg ccacggccca ggtgatgctg tccgacctgg cggcgcacat 4260 agcgatctgg tcctgaccc  
 gaccgaggcg atcaccaag ggtcgaaag ctgtccac 4320 gcgccaaac tgatgccgc gggcatccgc gacgacacgg  
 tgcgtctt tgcctggc 4380 cggccacgcgg atgacgtgat cgacggtcag gcccggca gcccggccga ggcgtgtac  
 4440 gacccgcagg cgcggctggc cggccgcgc gtcgacacgc tggccggcc gcaaggcgac 4500 ggtccggta cccggccctt  
 tgcggcgctg cgcgggtgg cgcggccggca tgatttcccg 4560 caggccgtgc ccatggacat gatcgaaggc ttgcgtatgg  
 atgtcgagggc ggcgcactat 4620 cgcacgctgg atgacgtgat ggaatattcc tatcacgtc caggcatcgt cggcgtgtac

4680 atggcccgcg tcatggcggt ggcgcacat cctgtccctgg accgcgcctg cgacactgggg 4740 ctggcggttcc agctgaccaa  
 catcgccgcg gacgtgatcg acgatgcgcg catcgccgcg 4800 tgctatctgc cgggggactg gctggaccag gcggggcgcgc  
 ggatcgacgg gccgggtgccc 4860 tcgcccggacgt tgtacacagt gatccctccgg ctgttggatg aggcggaaacc ctattacgcg  
 4920 tcggcgcggg tgggtctggc ggaatctgcca ccgcgcgtcg cctggtccat cgccgcgcg 4980 ctacggatct atcgccat  
 cgggctgcgc atccgcaaga gcggggccgca ggcctatcg 5040 cagcggatca gcacgtccaa ggctgcacaag atcggcctgc  
 tgggcgtcgg gggctggat 5100 gtcgcgcgt caccgcgtcc gggggcgggc gtgtcgccgc agggcctctg gacccggccg  
 5160 catcacgtct aggcgcgcgc ggcgttagggc agaaccgtt ccagcaggcgc cgcgatttc 5220 ggagcctgaa ggcgcgttgct  
 ggcgcgcgc gctccagg tggcgccgct ggcctcgtaa 5280 tgacggaca cgttctgcag gtctgacacg gccagaaggc  
 cgcgcgcgg gcccggggcc 5340 gcggcatcgc gaccggatc ctggccaagc gccgcctggc cgcccacgcgt gtccagcagg  
 5400 tcgtcatagg actggaaacac gcggcccagc tgacggccaa agtgcgtatcat ctgggtctgc 5460 tcctcgccgt cgaactccct  
 gatcacggcc agcatctcca gcccggcgt gaacagcagc 5520 ccggcttca ggtcctgttc ctgttcgacc cccgcgcgt  
 tcttggccgc gtgcagggtcc 5580 aggtcctggc cggcgcacag gcccgcggc cccagggacc gcgcacaggat ccgcaccagc  
 5640 tgcgcgcga ccgtgcccga cgcgcgcgc gcaccggcca gcagggccat tgcctcggt 5700 atcagggcga tgccgcggcag  
 cacggcacgg ctgtccat ggcgcacatg ggtcgccggc 5760 cggccgcggc gcagggccgc atcgccatg cagggcaggat  
 cgtcgaagat cagcgtatgcg 5820 gcatgcacca tctcgaccgc gcaggcggcg tcgacgtatg tgtcgcagac cccgcggcag  
 5880 gcctctgccc caagcagcat cagcatgcg cggaaaccggc tgcccgacga cagcgcgcga 5940 tggctcatgg ccgcgcggcag  
 cggctgcgc acggcaccga atccctgggc gatcttc 6000 agtctggtct gcagaagggt ggcgtggatc gggtttgacgt  
 ctctcat cagtgccttc 6060 ggcgttgggt tctgacccgttgg cggaaagggtc aggccggggc ggcacccgtt gacccgtcat  
 6120 ccaccgtcaa cagtcccat gttggaaacgg ttacgcgcgc attgcgagcc ttttcgacgg 6180 cgacgcgggg tcgcgcggca  
 atttgtccaa caagggtcagt gga 6223 <210> 2 <211> 729 <212> DNA <213> crtW  
 gene <400> 2 atgagcgcac atgcctgcgc caaggcagat ctgacccgtt ccagccgtat cgtctcggtt 60 ggcacatcg  
 cccgcgtggctt ggcgcgtat gtcgtatgcgc tgggttttgc ggacgcggcg 120 ggcacatccca tcctggcgat cgcgaatttc  
 ctggggcttga cctggctgtc ggtcggtctg 180 ttcttcatcg cgcgtatgcgc gatgcacggg tcgggtcgatc cggggcgttcc  
 ggcgcggcaat 240 ggcgcgtatgg ggcgttgggtt cctgtggctg tatgcggat tttcggttgc caagatgtatc 300  
 gtcaaggcaca tggccatca ccgcatacc ggaaccgcacg acgacccgtt tttcgaccat 360 ggcggccggg tcgcgtggta

cgccgcgttc atcggcacct atttcggctg gcgcgagggg 420 ctgctgctgc ccgtcatcgt gacggcttat gcgcgtatcc  
 tggggatcg ctggatgtac 480 gtggctitct ggccgctgcc gtcgatcctg gcgtcgatcc agctgttcgt gttcggcacc  
 540 tggctccgc accggcccccgg ccacgacgacg ttccggacc gcataatgc gcggtcgctcg 600 cggatcagcg accccgtgtc  
 gctgctgacc tgcttcaact ttggtggtta tcatacggaa 660 caccacgtc acccgacggt gccttggtgg cgccctgccc  
 gcacccgcac caagggggac 720 accgcatga  
 729 <210> 3 <211> 242 <212> PRT <213> crtW amino acid <400> 3 Met Ser Ala His Ala Leu Pro  
 Lys Ala Asp Leu Thr Ala Thr Ser Leu 1 5 10 15 Ile Val  
 Ser Gly Gly Ile Ile Ala Ala Trp Leu Ala Leu His Val His 20 25  
 30 Ala Leu Trp Phe Leu Asp Ala Ala His Pro Ile Leu Ala Ile Ala 35 40  
 45 Asn Phe Leu Gly Leu Thr Trp Leu Ser Val Gly Leu Phe Phe Ile Ala 50 55  
 60 His Asp Ala Met His Gly Ser Val Val Pro Gly Arg Pro Arg Gly Asn 65 70  
 75 80 Ala Ala Met Gly Gln Leu Val Leu Trp Leu Tyr Ala Gly Phe Ser Trp 8  
 90 95 Arg Lys Met Ile Val Lys His Met Ala His His Arg His Thr Gly Thr 100  
 105 110 Asp Asp Asp Pro Asp Phe Asp His Gly Gly Pro Val Arg Trp Tyr Ala 115  
 120 125 Arg Phe Ile Gly Thr Tyr Phe Gly Trp Arg Glu Gly Leu Leu Leu Pro 130  
 135 140 Val Ile Val Thr Val Tyr Ala Leu Ile Leu Gly Asp Arg Trp Met Tyr 145  
 150 155 160 Val Val Phe Trp Pro Leu Pro Ser Ile Leu Ala Ser Ile Gln Leu  
 Phe 165 170 175 Val Phe Gly Thr Trp Leu Pro His Arg Pro  
 Gly His Asp Ala Phe Pro 180 185 190 Asp Arg His Asn Ala Arg  
 Ser Ser Arg Ile Ser Asp Pro Val Ser Leu 195 200 205 Leu Thr Cys  
 Phe His Phe Gly Gly Tyr His His Glu His His Leu His 210 215 220 Pro  
 Thr Val Pro Trp Trp Arg Leu Pro Ser Thr Arg Thr Lys Gly Asp 225 230 235  
 240 Thr Ala <210> 4 <211> 489 <212> DNA <213> crtZ gene <400> 4 atgaccaatt tcctgatcgt  
 cgtcgccacc gtgctggta tggagttgac ggcctattcc 60 gtccaccgtt ggatcatgca cggcccccgt ggctggggct  
 ggcacaagtc ccaccacgag 120 gaacacgacc acgcgctgga aaagaacgac ctgtacggcc tggctttgc ggtgatcgcc

180 acgggtctgt tcacgggtggg ctggatctgg gcgcgggtcc tgtggtggat cgcttgggc 240 atgaccgtct atgggctgat  
 ctatttcgtc ctgcgtacg ggctgggtca tcagcgctgg 300 ccgttccgct atatcccgctg caagggtat gcccggcc  
 tgtatcaggc ccaccgcctg 360 caccacgcgg tcgaggggacg cgaccattgc gtcagcttcg gcttcata tgcggccgg  
 420 gtcgacaagc tgaaggcagga cctgaagacg tcggggtgc tgcggccga ggccgcaggag 480 cgcacgtga  
 489 <210> 5 <211> 162 <212> PRT <213> crtZ amino acid <400> 5 Met Thr Asn Phe Leu Ile Val  
 Val Ala Thr Val Leu Val Met Glu Leu 1 5 10 15 Thr Ala  
 Tyr Ser Val His Arg Trp Ile Met His Gly Pro Leu Gly Trp 20 25  
 30 Gly Trp His Lys Ser His His Glu Glu His Asp His Ala Leu Glu Lys 35 40  
 45 Asn Asp Leu Tyr Gly Leu Val Phe Ala Val Ile Ala Thr Val Leu Phe 50 55  
 60 Thr Val Gly Trp Ile Trp Ala Pro Val Leu Trp Trp Ile Ala Leu Gly 65 70  
 75 80 Met Thr Val Tyr Gly Leu Ile Tyr Phe Val Leu His Asp Gly Leu Val 8  
 90 95 His Gln Arg Trp Pro Phe Arg Tyr Ile Pro Arg Lys Gly Tyr Ala Arg 100  
 105 110 Arg Leu Tyr Gln Ala His Arg Leu His His Ala Val Glu Gly Arg Asp 115  
 120 125 His Cys Val Ser Phe Gly Phe Ile Tyr Ala Pro Pro Val Asp Lys Leu 130  
 135 140 Lys Gln Asp Leu Lys Thr Ser Gly Val Leu Arg Ala Glu Ala Gln Glu 145  
 150 155 160 Arg Thr <210> 6 <211> 1161 <212> DNA <213> crtY  
 gene <400> 6 gtgacccatg acgtgctgct ggcaggggcg ggccttgcga acgggtatcgccctggc 60 ctgcgcgcgg  
 cgccggccga cctgccccgtc ctgctgtgg atcatgcggc gggaccgtca 120 gacggccata cctggccctg ccacgacccc  
 gatctgtcgc cgcactggct ggccggcgtg 180 aagccctgc gcccgcctaa ctggcccgac caggagggtgc gctttcccg  
 ccatgcccgg 240 cggctggcca ccgggtacgg gtcgtggac gggccggcgc tggcggatgc ggtggccgg 300  
 tcggccggc agatccgctg gaacagcgcac atcgccctgc tggatgaaca gggggcgcacg 360 ctgtcctgcg gcacccggat  
 cgaggcggc gcggtccctgg acgggcgcgg cgccgcggc 420 tcgcggcatc tgaccgtggg tttccagaaa ttctgtggcgc  
 tcgagatcga gaccgactgc 480 cccccacggcg tgccccccc gatgatcatg gacgcgcaccc tcacccagca ggacgggtac  
 540 cgattcatct atctgctgcc cttctctccg acgcgcattcc tgatcgagga cactcgat 600 tccgatggcg gcaatctgga  
 cgacgacgcg ctggccggcgg cgtcccacga ctatgcccgc 660 cagcagggtcg ggaccggggc cgagggtccgg cgccgaacgcg

gcacccatgcctt cattgcgtg . . . . . 720 gccccatgacg cggcgggctt ctggccat cacgcggagg ggcctgttcc cgtgggactg  
 780 cgcgccgggt tcttcaccc ggtcaccggc tattcgctgc cctatgcggc gcaggtggc . . . . . 840 gacgtggtgg cggccctgc  
 cggggccccc ggcaccgacg cgctgcgcgg cgccatccgc . . . . . 900 gattacgcga tcgaccgggc acgcccgtgac cgcttctgc  
 gcctgctgaa ccggatgctg . . . . . 960 ttccgcggct gcgcgcggc ccggcgctat accctgctgc agcggttcta ccgcattgcgc  
 1020 catggactga tcgaacggtt ctatgcggc cggctgagcg tggcggatca gctgcgcac . . . . . 1080 gtgaccggca agccctccat  
 tccccctggc acggccatcc gctgcctgcc cgaacgtccc . . . . . 1140 ctgctgaagg aaaacgcatt a  
 1161 <210> 7 <211> 386 <212> PRT <213> crtY amino acid <400> 7 Val Thr His Asp Val Leu Leu  
 Ala Gly Ala Gly Leu Ala Asn Gly Leu 1 5 10 15 Ile Ala -  
 Leu Ala Leu Arg Ala Ala Arg Pro Asp Leu Arg Val Leu Leu 20 25  
 30 Leu Asp His Ala Ala Gly Pro Ser Asp Gly His Thr Trp Ser Cys His 35 40  
 45 Asp Pro Asp Leu Ser Pro His Trp Leu Ala Arg Leu Lys Pro Leu Arg 50 55  
 60 Arg Ala Asn Trp Pro Asp Gln Glu Val Arg Phe Pro Arg His Ala Arg 65 70  
 75 80 Arg Leu Ala Thr Gly Tyr Gly Ser Leu Asp Gly Ala Ala Leu Ala Asp 8  
 90 95 Ala Val Ala Arg Ser Gly Ala Glu Ile Arg Trp Asn Ser Asp Ile Ala 100  
 105 110 Leu Leu Asp Glu Gln Gly Ala Thr Leu Ser Cys Gly Thr Arg Ile Glu 115  
 120 125 Ala Gly Ala Val Leu Asp Gly Arg Gly Ala Gln Pro Ser Arg His Leu 130  
 135 140 Thr Val Gly Phe Gln Lys Phe Val Gly Val Glu Ile Glu Thr Asp Cys 145  
 150 155 160 Pro His Gly Val Pro Arg Pro Met Ile Met Asp Ala Thr Val Thr  
 Gln 165 170 175 Gln Asp Gly Tyr Arg Phe Ile Tyr Leu Leu  
 Pro Phe Ser Pro Thr Arg 180 185 190 Ile Leu Ile Glu Asp Thr  
 Arg Tyr Ser Asp Gly Gly Asn Leu Asp Asp 195 200 205 Asp Ala Leu  
 Ala Ala Ala Ser His Asp Tyr Ala Arg Gln Gln Gly Trp 210 215 220 Thr  
 Gly Ala Glu Val Arg Arg Glu Arg Gly Ile Leu Pro Ile Ala Leu 225 230 235  
 240 Ala His Asp Ala Ala Gly Phe Trp Ala Asp His Ala Glu Gly Pro Val 245  
 250 255 Pro Val Gly Leu Arg Ala Gly Phe Phe His Pro Val Thr Gly Tyr Ser 260

265 270 Leu Pro Tyr Ala Ala Gln Val Ala Asp Val Val Ala Gly Leu Ser Gly 275  
 280 285 Pro Pro Gly Thr Asp Ala Leu Arg Gly Ala Ile Arg Asp Tyr Ala Ile 290  
 295 300 Asp Arg Ala Arg Arg Asp Arg Phe Leu Arg Leu Leu Asn Arg Met Leu 305  
 310 315 320 Phe Arg Gly Cys Ala Pro Asp Arg Arg Tyr Thr Leu Leu Gln Arg  
 Phe 325 330 335 Tyr Arg Met Pro His Gly Leu Ile Glu Arg  
 Phe Tyr Ala Gly Arg Leu 340 345 350 Ser Val Ala Asp Gln Leu  
 Arg Ile Val Thr Gly Lys Pro Pro Ile Pro 355 360 365 Leu Gly Thr  
 Ala Ile Arg Cys Leu Pro Glu Arg Pro Leu Leu Lys Glu 370 375 380 Asn  
 Ala 385 <210> 8 <211> 1506 <212> DNA <213> crtI gene <400> 8 atgaacgccc attcgccgc  
 ggccaagacc gccatcgta tcggcgcagg cttggcgagg 60 ctggccctgg ccatccgcgc gcagtcgcgc ggcatcgcca  
 ccaccctggc cgaggccgg 120 gacaagcccg gcgggcgcgc ctatgtctgg cacgatcagg gccatgtctt cgacgcggc  
 180 ccgaccgtca tcaccgaccc cgatgcgcgc aaggagctgt gggcgctgac cgggcaggac 240 atggcgcgc acgtgacgct  
 gatgccggtg tcgcccctct atcgactgat gtggccggc 300 gggaaaggct tcgattacgt gaacgaggcc gatcagctgg  
 agcgcagat cgcccgatcc 360 aacccggacg acctggaagg ataccgcgc ttccgtgatt acgcggagga ggtgtatcag  
 420 gagggtacg tcaagctggg caccgtgcc ttcctcaagc tggggccagat gctcaaggcc 480 gcgccgcgc tgatgaagct  
 ggaggcctat aagtccgtcc atgccaaggat cgcgaccatcc 540 atcaaggacc cctatctgcgc gcaggcgat tcgtatcaca  
 cgctgctggt gggcgggaaat 600 cccttctgca ccagctcgat ctatgcgcgt atccacgcgc tggagcggc cggcggggtc  
 660 tggttcgcca agggcggcac caaccagctg gtcgcggca tggtcgcgt gttcgaacgg 720 cttggcggcc agatgtatgc  
 gaacgcgaag gtcgcccggta tcgagaccga gggcgccgg 780 accacggcgc tcaccctggc ggacggcgg tcttaaggg  
 ccgacatggt cgcccgac 840 ggcgacgtca tgcacaacta tcgcgcacctg ctggccaca cggccgcgg gcagagccgc  
 900 gcgaaatcgc tggaccgcaa ggcgtggcc atgtcgatgt tcgtgcgtca ttccggatctg 960 cgcgaggcgc ccaaggacat  
 cgccatcac accatccgt tcggcccccgt ctacaggag 1020 ctggtcaacg agatcttcaa gggcccgaag ctggccggagg  
 atttctcgat gtacatcgat 1080 tcggccctgca cgaccatcc ggacatggcgc ctcggccaca tgcacgcgtca ttacgtgcgt  
 1140 gccccgtgc cgcatctggg cccgcggcgg atcgattggg cggtcgaggg gcccgcgtat 1200 gcccggca tcctggcgtc  
 cctggaggag cggctgatcc cgaacctgca cggccaaacctg 1260 accacgacgc gcatcttcaac gcccggat ttcgcccagcg

aactgaacgc ccatcacggc 1320 agcgccitct cggtcgagcc gatcctgacg caatccgcgt gttccggcc gcacaaccgc  
 1380 gacaagacga tccgcaactt ctatctggtc ggcgcggca cccatccggg cgccggcatt 1440 cggggcgctcg tgggctcgcc  
 caaggccacg gcccagggtga tgctgtccga cctggcggc 1500 gcatga  
 1506 <210> 9 <211> 501 <212> PRT <213> crtI amino acid <400> 9 Met Asn Ala His Ser Pro Ala  
 Ala Lys Thr Ala Ile Val Ile Gly Ala 1 5 10 15 Gly Phe  
 Gly Gly Leu Ala Leu Ala Ile Arg Leu Gln Ser Ala Gly Ile 20 25  
 30 Ala Thr Thr Leu Val Glu Ala Arg Asp Lys Pro Gly Gly Arg Ala Tyr 35 40  
 45 Val Trp His Asp Gln Gly His Val Phe Asp Ala Gly Pro Thr Val Ile 50 55  
 60 Thr Asp Pro Asp Ala Leu Lys Glu Leu Trp Ala Leu Thr Gly Gln Asp 65 70  
 75 80 Met Ala Arg Asp Val Thr Leu Met Pro Val Ser Pro Phe Tyr Arg Leu 8  
 90 95 Met Trp Pro Gly Gly Lys Val Phe Asp Tyr Val Asn Glu Ala Asp Gln 100  
 105 110 Leu Glu Arg Gln Ile Ala Gln Phe Asn Pro Asp Asp Leu Glu Gly Tyr 115  
 120 125 Arg Arg Phe Arg Asp Tyr Ala Glu Glu Val Tyr Gln Glu Gly Tyr Val 130  
 135 140 Lys Leu Gly Thr Val Pro Phe Leu Lys Leu Gly Gln Met Leu Lys Ala 145  
 150 155 160 Ala Pro Ala Leu Met Lys Leu Glu Ala Tyr Lys Ser Val His Ala  
 Lys 165 170 175 Val Ala Thr Phe Ile Lys Asp Pro Tyr Leu  
 Arg Gln Ala Phe Ser Tyr 180 185 190 His Thr Leu Leu Val Gly  
 Gly Asn Pro Phe Ser Thr Ser Ile Tyr 195 200 205 Ala Leu Ile  
 His Ala Leu Glu Arg Arg Gly Gly Val Trp Phe Ala Lys 210 215 220 Gly  
 Gly Thr Asn Gln Leu Val Ala Gly Met Val Ala Leu Phe Glu Arg 225 230 235  
 240 Leu Gly Gly Gln Met Met Leu Asn Ala Lys Val Ala Arg Ile Glu Thr 245  
 250 255 Glu Gly Ala Arg Thr Thr Gly Val Thr Leu Ala Asp Gly Arg Ser Leu 260  
 265 270 Arg Ala Asp Met Val Ala Ser Asn Gly Asp Val Met His Asn Tyr Arg 275  
 280 285 Asp Leu Leu Gly His Thr Ala Arg Gly Gln Ser Arg Ala Lys Ser Leu 290  
 295 300 Asp Arg Lys Arg Trp Ser Met Ser Leu Phe Val Leu His Phe Gly Leu 305

310 . . . . . 315 . . . . . 320 Arg Glu Ala Pro Lys Asp Ile Ala His His Thr Ile Leu Phe Gly  
 Pro . . . . . 325 . . . . . 330 . . . . . 335 Arg Tyr Arg Glu Leu Val Asn Glu Ile Phe  
 Lys Gly Pro Lys Leu Ala . . . . . 340 . . . . . 345 . . . . . 350 Glu Asp Phe Ser Leu Tyr  
 Leu His Ser Pro Cys Thr Thr Asp Pro Asp . . . . . 355 . . . . . 360 . . . . . 365 Met Ala Pro  
 Pro Gly Met Ser Thr His Tyr Val Leu Ala Pro Val Pro . . . . . 370 . . . . . 375 . . . . . 380 His  
 Leu Gly Arg Ala Glu Ile Asp Trp Ala Val Glu Gly Pro Arg Tyr 385 . . . . . 390 . . . . . 395  
 400 Ala Asp Arg Ile Leu Ala Ser Leu Glu Glu Arg Leu Ile Pro Asn Leu . . . . . 405  
 410 . . . . . 415 Arg Ala Asn Leu Thr Thr Arg Ile Phe Thr Pro Ala Asp Phe Ala . . . . . 420  
 425 . . . . . 430 Ser Glu Leu Asn Ala His His Gly Ser Ala Phe Ser Val Glu Pro Ile . . . . . 435  
 440 . . . . . 445 Leu Thr Gln Ser Ala Trp Phe Arg Pro His Asn Arg Asp Lys Thr Ile . . . . . 450  
 455 . . . . . 460 Arg Asn Phe Tyr Leu Val Gly Ala Gly Thr His Pro Gly Ala Gly Ile 465  
 470 . . . . . 475 . . . . . 480 Pro Gly Val Val Gly Ser Ala Lys Ala Thr Ala Gln Val Met Leu  
 Ser . . . . . 485 . . . . . 490 . . . . . 495 Asp Leu Ala Gly Ala . . . . . 500 <  
 210> 10 <211> 915 <212> DNA <213> crtB gene <400> 10 atgagcgatc tggcctgac ctgcaccgag  
 gcgatcaccc aagggtcgca aagcttgc . . . . . 60 acggcgccca agctgatgcc gcccggcatc cgcgacgaca cggtgatgtc  
 ctatgcctgg . . . . . 120 tgccgccacg cggatgacgt gatcgacgtt caggccctgg gcagccgccc cgaggcggtg . . . . . 180  
 aacgaccgcg aggccggct ggacggcccg cgcgtcgaca cgctggccgc cctgcagg . . . . . 240 gacggtccgg tgaccccgcc  
 ctttgcgcgc ctgcgcgcgg tggcgcggcg gcatgatttc . . . . . 300 ccgcaggcct ggcccatgga cctgatcgaa ggcttcgcga  
 tggatgtcga ggccgcgcac . . . . . 360 tatgcacgc tggatgacgt gctgaaatat tcctatcacg tcgcaggcat cgccggcgt  
 420 atgatggccc gcgtgtatggg cgtgcgcgac gatccgtcc tggaccgcgc ctgcgacctg . . . . . 480 gggctggcgt tccagctgac  
 caacatcgcg cgcgacgtga tcgacgatgc ggcgcac . . . . . 540 cggtgctatc tgccggggga ctggctggac caggccggcg  
 cggatcga cggccgggtg . . . . . 600 ccgtcgccgg agctgtacac agtgcacgc cggctgttgg atgaggcggaa accctattac  
 660 gcgtcgccgc ggggtgggtct ggcggatctg ccaccgcgt ggcgcgttgc catgcgcgc . . . . . 720 ggcgtacgga tctatcgcc  
 catcgggcgt cgcacccgca agagccggcc gcaggccat . . . . . 780 cggccacgc tcagcagtc caaggctgcc aagatccggc  
 tgctgggcgt cggggcgtgg . . . . . 840 gatgtcgccgac gatcgcgcct gcccggggcg ggcgtgtcgc ggcaggccct ctggacccgg

900 ccgcacg tctag 915 <210> 11 <211> 304

<212> PRT <213> crtB amino acid <400> 11 Met Ser Asp Leu Val Leu Thr Ser Thr Glu Ala Ile Thr Gln

Gly Ser 1	5	10	15	Gln Ser Phe Ala Thr Ala Ala Lys Leu
Met Pro Pro Gly Ile Arg Asp	20	25	30	Asp Thr Val Met Leu
Tyr Ala Trp Cys Arg His Ala Asp Asp Val Ile	35	40	45	Asp Gly
Gln Ala Leu Gly Ser Arg Pro Glu Ala Val Asn Asp Pro Gln	50	55	60	
Ala Arg Leu Asp Gly Leu Arg Val Asp Thr Leu Ala Ala Leu Gln Gly	65	70	75	
80 Asp Gly Pro Val Thr Pro Pro Phe Ala Ala Leu Arg Ala Val Ala Arg		85	90	
95 Arg His Asp Phe Pro Gln Ala Trp Pro Met Asp Leu Ile Glu Gly Phe		100	105	
110 Ala Met Asp Val Glu Ala Arg Asp Tyr Arg Thr Leu Asp Asp Val Leu		115	120	
125 Glu Tyr Ser Tyr His Val Ala Gly Ile Val Gly Val Met Met Ala Arg		130	135	
140 Val Met Gly Val Arg Asp Asp Pro Val Leu Asp Arg Ala Cys Asp Leu		145	150	
155	160 Gly Leu Ala Phe Gln Leu Thr Asn Ile Ala Arg Asp Val Ile Asp Asp			
165	170	175 Ala Arg Ile Gly Arg Cys Tyr Leu Pro Gly Asp Trp Leu Asp Gln		
Ala	180	185	190 Gly Ala Arg Ile Asp Gly Pro Val Pro Ser Pro	
Glu Leu Tyr Thr Val	195	200	205 Ile Leu Arg Leu Leu Asp Glu Ala	
Glu Pro Tyr Tyr Ala Ser Ala Arg	210	215	220 Val Gly Leu Ala Asp Leu	
Pro Pro Arg Cys Ala Trp Ser Ile Ala Ala	225	230	235	240
Ala Leu Arg Ile Tyr Arg Ala Ile Gly Leu Arg Ile Arg Lys Ser Gly			245	250
255 Pro Gln Ala Tyr Arg Gln Arg Ile Ser Thr Ser Lys Ala Ala Lys Ile		260	265	
270 Gly Leu Leu Gly Val Gly Gly Trp Asp Val Ala Arg Ser Arg Leu Pro		275	280	
285 Gly Ala Gly Val Ser Arg Gln Gly Leu Trp Thr Arg Pro His His Val		290	295	
300 <210> 12 <211> 882 <212> DNA <213> crtE gene <400>	12 atgagacgag acgtcaaccc gatccacgc			
acccttctgc agaccagact tgaggagatc	60 gcccaggat tcggtgccgt gtcgcagccg ctcggcgcgg ccatgagcc			
tggcgctgt	120 tcgtcggca ggcggttccg cggcatgctg atgctgcttg cggcagaggc ctcgggcggg	180		

gtctgcgaca cgatcgatcgac cgccgcctgc gcggtcgaga tggtcgtgc cgcacatcgctg 240 atcttcgacg acctggccctg  
 catggacgat gcccggctgc gcccggccg gcccgcgacc 300 catgtggcgc atggcgaaag ccgtggccgtg ctggggcggca  
 tcggccctgat caccgaggca 360 atggccctgc tggccgggtgc gcccggccg tcgggcacgg tgcggccgca gctgggtgcgg  
 420 atcctgtcgc ggtccctggg gccgcaggcc ctgtgcgcg gcccaggacct ggacctgcac 480 gcggccaaga acggcgcggg  
 ggtcgaaacag gaacaggacc tgaagaccgg cgtgctgttc 540 atcgcggggc tggagatgct ggccgtgatc aaggagttcg  
 acgcccggaggaa gcagaccagg 600 atgatcgact ttggccgtca gctggccgc gtgttccagt cctatgacga cctgctggac  
 660 gtcgtggcgc accaggccgc gcttggcaag gataccggtc gcgatgcccgc ggccccccgc 720 ccgcggccgc gccttcggc  
 cgtgtcagac ctgcagaacg tgcgtccgtca ttacgaggcc 780 agccgcgccc aactggacgc gatgctgcgc agcaagcgc  
 ttccaggctcc ggaaatcgcg 840 gcccctgtgg aacgggttct gcctacgccc gcgcgcgcct ag  
 882 <210> 13 <211> 293 <212> PRT <213> crtE amino acid <400> 13 Met Arg Arg Asp Val Asn Pr  
 Ile His Ala Thr Leu Leu Gln Thr Arg 1 5 10 15 Leu Glu  
 Glu Ile Ala Gln Gly Phe Gly Ala Val Ser Gln Pro Leu Gly 20 25  
 30 Ala Ala Met Ser His Gly Ala Leu Ser Ser Gly Arg Arg Phe Arg Gly 35 40  
 45 Met Leu Met Leu Leu Ala Ala Glu Ala Ser Gly Gly Val Cys Asp Thr 50 55  
 60 Ile Val Asp Ala Ala Cys Ala Val Glu Met Val His Ala Ala Ser Leu 65 70  
 75 80 Ile Phe Asp Asp Leu Pro Cys Met Asp Asp Ala Gly Leu Arg Arg Gly 8  
 90 95 Arg Pro Ala Thr His Val Ala His Gly Glu Ser Arg Ala Val Leu Gly 100  
 105 110 Gly Ile Ala Leu Ile Thr Glu Ala Met Ala Leu Leu Ala Gly Ala Arg 115  
 120 125 Gly Ala Ser Gly Thr Val Arg Ala Gln Leu Val Arg Ile Leu Ser Arg 130  
 135 140 Ser Leu Gly Pro Gln Gly Leu Cys Ala Gly Gln Asp Leu Asp Leu His 145  
 150 155 160 Ala Ala Lys Asn Gly Ala Gly Val Glu Gln Glu Gln Asp Leu Lys  
 Thr 165 170 175 Gly Val Leu Phe Ile Ala Gly Leu Glu Met  
 Leu Ala Val Ile Lys Glu 180 185 190 Phe Asp Ala Glu Glu Gln  
 Thr Gln Met Ile Asp Phe Gly Arg Gln Leu 195 200 205 Gly Arg Val  
 Phe Gln Ser Tyr Asp Asp Leu Leu Asp Val Val Gly Asp 210 215 220 Gln

Ala Ala Leu Gly Lys Asp Thr Gly Arg Asp Ala Ala Ala Pro Gly 225 230 235  
240 Pro Arg Arg Gly Leu Leu Ala Val Ser Asp Leu Gln Asn Val Ser Arg 245  
250 255 His Tyr Glu Ala Ser Arg Ala Gln Leu Asp Ala Met Leu Arg Ser Lys 260  
265 270 Arg Leu Gln Ala Pro Glu Ile Ala Ala Leu Leu Glu Arg Val Leu Pro 275  
280 285 Tyr Ala Ala Arg Ala 290 <210> 14 <211> 19 <212> DNA <213>  
Artificial Sequence <220> <223> forward primer for crt gene <400> 14 gttccacgac tggggcatac  
19 <210> 15 <211> 28 <212> DNA <213> Artificial Sequence <220> <223> reverse primer for cr  
gene <400> 15 tccactgacc ttgttggaca aattgccg 28